Optical Spectroscopy and Velocity Dispersions of SZ-Selected Galaxy Clusters

Optical Spectroscopy and Velocity Dispersions of SZ-selected Galaxy Clusters

A dissertation presented

by

Jonathan Ruel

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

May 2013

Thesis advisor

Christopher W. Stubbs

Author

Jonathan Ruel

Optical Spectroscopy and Velocity Dispersions of SZ-selected Galaxy Clusters

Abstract

We present the design and the first results of a program of optical spectroscopy of galaxies in clusters detected in South Pole Telescope (SPT) data using the Sunyaev-Zel'dovich (SZ) effect, the spectral distortion of the cosmic microwave background from galaxy clusters. We use resampling for an empirical determination of the uncertainty in cluster velocity dispersion calculated from galaxy redshift measurements. We discuss outstanding questions that need answering in order to reach the goal of using cluster velocity dispersion measurements to calibrate scaling relation between the SZ observable and the cluster mass.

Contents

Title Page i
Copyright Page ii
Abstract iii
Table of Contents iv
Relation to Previously Published Work vi
Acknowledgments vii
Dedication viii
Epigraph ix

1 Introduction and summary **1**
1.1 Dark matter, the growth of structure, and cosmology with clusters . . 3
1.2 The hot gas 8
1.2.1 The Sunyaev-Zel'dovich effect 8
1.2.2 The South Pole Telescope cluster survey 11
1.3 Galaxies 14
1.3.1 Dynamical properties 15
1.4 Structure of this thesis 17

2 Optical spectroscopy and multi-wavelength follow-up of SPT clusters **19**
2.1 The place of optical spectroscopy 19
2.2 Combining multiple mass observables 21
2.2.1 Mass bias from scatter and cluster selection 22
2.2.2 Follow-up masses for individual clusters 25
2.2.3 Combining multiple mass observables 31
Appendix: multiple scaling relation statistics 34

3 A spectroscopic follow-up program for the SPT-SZ cluster survey **46**
3.1 A few-$N_{\rm members}$ spectroscopic strategy 46
3.2 Observations: South Pole Telescope 51
3.3 Observations: optical spectroscopy 54

4 Spectroscopic data processing **64**
4.1 CCD reductions . . . 65
4.2 Extracted spectrum processing and cross-correlation for velocity . . . 75

5 Methods and statistics of velocity dispersion measurements **80**
5.1 Standard methods . . . 81
5.1.1 From redshifts to peculiar velocities . . . 81
5.1.2 Dispersion and velocity errors . . . 82
5.1.3 Phase-space membership selection . . . 84
5.1.4 Resistant and robust estimators . . . 89
5.2 The stacked cluster . . . 91
5.3 Statistical methodology in the few-$N_{\rm members}$ regime . . . 95
5.3.1 Generated subsamples . . . 95
5.3.2 Unbiased estimators . . . 96
5.3.3 Confidence intervals . . . 102
5.3.4 Impact of membership selection . . . 108

6 Data – optical spectroscopy of 60 SPT clusters **110**
6.1 Results . . . 111
6.1.1 Individual galaxy redshifts . . . 111
6.1.2 Cluster redshifts and velocity dispersions . . . 111
6.1.3 Central galaxy peculiar velocities . . . 116
6.1.4 Data in the literature: summary and comparison . . . 124
6.2 Comparison of velocity dispersions with other observables . . . 125
6.2.1 Comparison with SZ-based masses . . . 125
6.2.2 Comparison with X-ray observations . . . 129

7 Conclusion and next steps **135**

Relation to Previously Published Work

Large portions of Chapters 3 to 6 are set to appear in the following paper:

> "Optical Spectroscopy and Velocity Dispersions of Galaxy Clusters from the South Pole Telescope Survey", J. Ruel, G. Bazin, M. Bayliss, M. Brodwin, R. J. Foley, B. Stalder, *et al.*, in prep.

Partial results included here were previously published with a different emphasis in

> "SPT-CL J0546-5345: A Massive $z > 1$ Galaxy Cluster Selected Via the Sunyaev-Zel'dovich Effect with the South Pole Telescope", M. Brodwin, J. Ruel, *et al. Ap. J.* **721**, 2010
>
> "Discovery and Cosmological Implications of SPT-CL J2106-5844, the Most Massive Known Cluster at $z > 1$", R. J Foley, K. Andersson, G. Bazin, T. de Haan, J, Ruel, *et al. Ap. J.* **731**, 2011
>
> "SPT-CL J0205-5829: A $z = 1.32$ Evolved Massive Galaxy Cluster in the South Pole Telescope Sunyaev-Zel'dovich Effect Survey", B. Stalder, J. Ruel, R. Suhada, M. Brodwin, *et al. Ap. J.* **763**, 2013
>
> "A massive, cooling-flow-induced starburst in the core of a luminous cluster of galaxies", M. McDonald, M. Bayliss, B. A. Benson, R. J. Foley, J. Ruel, P. Sullivan, S. Veilleux, *et al. Nature* **488**, 2012

Acknowledgments

I would like to thank Christopher Stubbs for giving me the opportunity to work on this research project, as well as for his constant support and advice over the years.

I also wish to express gratitude to the other members of my thesis committee, Douglas Finkbeiner and John Doyle.

The completion of this work would have been impossible without the teaching, support and friendship of the colleagues with whom I worked directly, most of all Ryan Foley, but also Brian Stalder, Mark Brodwin, Matt Bayliss, Gurvan Bazin, Armin Rest and Will High. I also credit the help and advice of many other members of the SPT team, most notably Mike Gladders, Mike McDonald, and Tony Stark.

Thanks to Gautham Narayan, Daniyar Nurgaliev, Amali Vaz and Peter Doherty for their friendship and support, and for sharing information regarding the location of our Dear Leader.

Infinite recognition to my wife, Katie Rose, for everything.

Merci à mes parents pour leur soutien soutenu dès le début de mes études.

I cannot dream of ever listing all the Harvard staff members, students and faculty who were important friends during my time in graduate school. I would like to extend special thanks to Susan Zawalich and Jim and Doreen Hogle of Dudley House, and Bari Walsh of GSAS publications.

Dedicated to Katie
you made everything possible

Sur la terre délaissée mes souvenirs ont poussé
Nourris par les rayons du soleil et l'odeur du fumier
–Pierre Lapointe (Hyacinthe la jolie)

Chapter 1

Introduction and summary

The present chapter provides essential background information about the physical properties of clusters of galaxies and their relation to cosmological studies, before summarizing the structure and results of this thesis.

When observed at optical wavelengths, a cluster of galaxies appears as a gravitationally bound group of galaxies. For instance, Figure 1.1 shows a false-color image of SPT-CL J0438-5419 (Williamson et al., 2011), a massive galaxy cluster at $z = 0.422$. The large, diffuse galaxy at the center and the objects surrounding it which have the same yellow color are galaxies associated with the cluster. Large — or "rich" — galaxy clusters have hundreds of member galaxies, and the possible number of members for a cluster spans the whole range all the way down to a handful of galaxies. Those small — or "poor" — clusters are often called "groups", although no discrete physical distinction exists between groups and larger clusters.

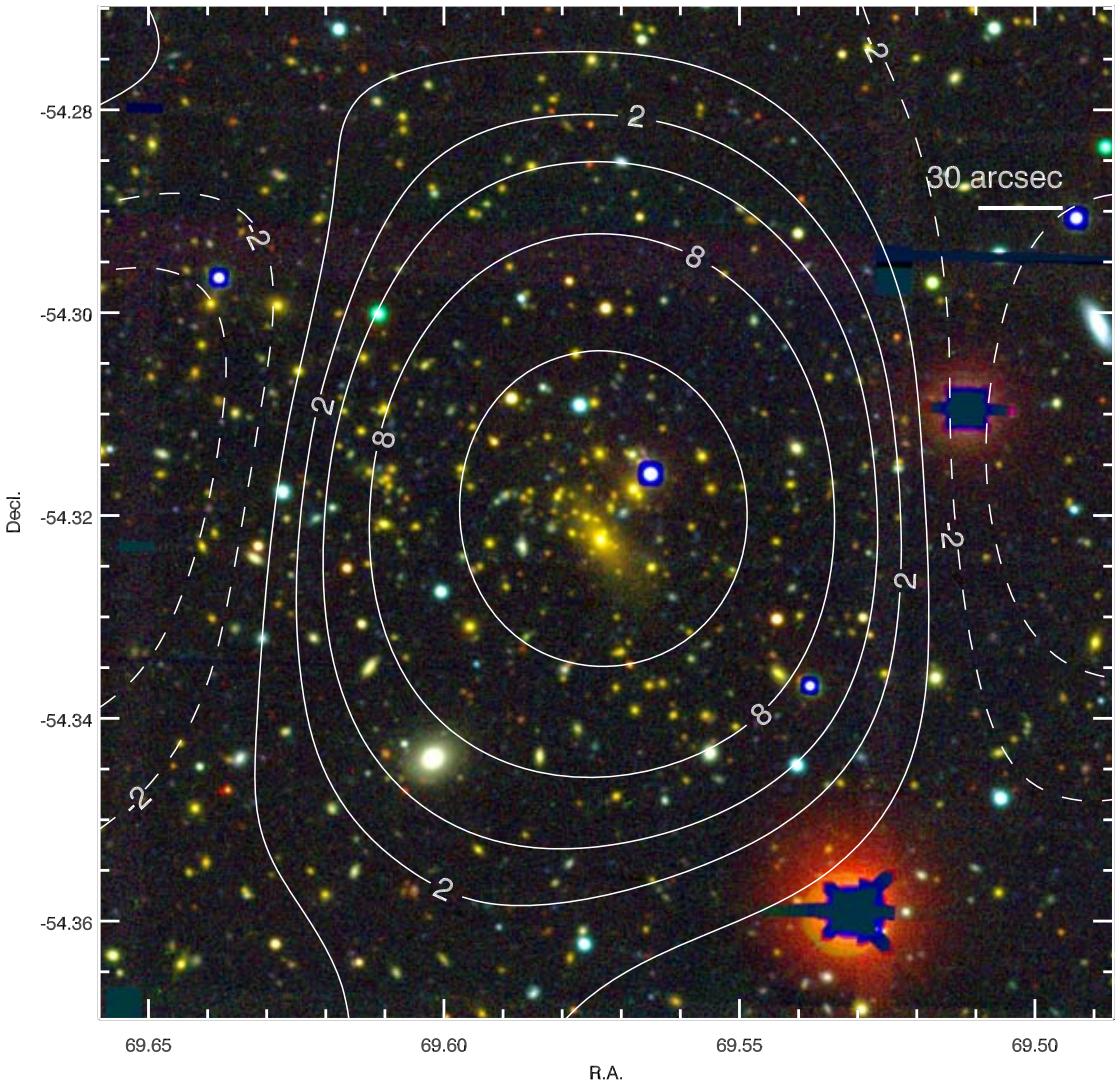

Figure 1.1: Optical image and SZ contours of SPT-CL J0438-5419 (Williamson et al., 2011), a massive South Pole Telescope (SPT) cluster at $z = 0.422$. The R, G, and B channels are respectively i-, r- and g-filter images taken with MOSAIC-II on the CTIO Blanco 4-meter telescope. The contours show the SZ decrement signal-to-noise ratio, as observed in the SPT CMB observations.
Image credit: adapted from Williamson et al. (2011).

Clusters were first described – and named – by the observation of their galaxies; see Biviano (2000) for a history of early optical cluster studies. It is now well established that most of the baryonic matter in a cluster resides not in its galaxies but in a

hot, large halo (or large enveloping volume) of intracluster gas. The mass budget of the cluster is in turn dominated by dark matter, which accounts for about 85% of a cluster's mass. We will look at those different components in a logical, if ahistorical order, starting with the dark matter, then moving on to the gas, and finishing with the galaxies. Voit (2005) and Allen et al. (2011) are good recent reviews of galaxy cluster properties and cosmological implications.

1.1 Dark matter, the growth of structure, and cosmology with clusters

The picture of the evolution of large-scale structure and the formation of clusters of galaxies in the ΛCDM concordance cosmology is that the matter distribution in the early universe was very smooth and its thermal state homogeneous, as evidenced by the Cosmic Microwave Background (CMB). There were small density perturbations, seen as temperature fluctuations of order 10^{-5} in the CMB. With time, these overdense regions accreted more matter and grew via gravity; the largest overdensity regions evolved into clusters of galaxies. This collapse process is dominated by dark matter and has been studied in dark-matter (gravity only) N-body simulations.

Defining mass

The overdensities or clusters do not have a discrete physical boundary, therefore there are different ways to define their characteristic mass and radius. We will define $R_{\Delta c}$ as the radius from the cluster center within which the average density is Δ

times the critical density (hence the subscript c) at the cluster redshift, $\rho_{crit}(z) = 3H^2(z)/8\pi G$, where $H(z)$ is the Hubble expansion parameter. Then we define $M_{\Delta c}$ as the mass contained within $R_{\Delta c}$. R_{500c} and M_{500c} are most commonly used in the literature for gas-based mass measurements, like those based on X-ray or Sunyaev-Zel'dovich (SZ) observations (see Section 1.2). R_{200c} and M_{200c} are more appropriate for the radius probed by optical studies, velocity dispersions and weak lensing. The latter is closer to, but still smaller than the "virial radius", the radius within which the matter is in equilibrium with the potential and its kinetic and potential energy are related via the so-called virial theorem. The virial radius can be calculated to be R_{178c} in a spherical top-hat collapse model. Many publications call R_{200c} the "virial radius"[1].

The cluster mass function

The abundance of galaxy clusters is described by the *cluster mass function* (Press & Schechter, 1974), which we will write $n_M(M, z)$; it gives the number of overdensities with a mass greater than M per unit volume at redshift z. It can be used to construct observables such as the number of clusters in an observable cosmological volume above a given mass. The cluster mass function is dependent on cosmology, and therefore its measurement can be used to derive empirical constraints on cosmological parameters. As far as dark energy is concerned, this can be understood in the following way: clusters grow by accretion through gravity; dark energy counteracts gravity on very

[1] An alternative definition of the overdensity radius sometimes used in the literature is with respect to the *mean* or *background* density (called either $R_{\Delta b}$ or $R_{\Delta m}$). An alternative definition of "virial radius" is R_{180m}; $R_{200c} < R_{178c} < R_{180m}$ but their values are similar. Many publications omit the subscripts b, m or c altogether.

large scales; therefore the accretion will happen at different rates depending on the properties of dark energy.

This overdensity growth process and the cluster mass function have been studied in cosmological N-body simulations of dark matter particles, notably in Tinker et al. (2008). Tinker et al. (2008) parametrize the cluster mass function in such a way that it is a factor that depends on cosmology (i.e., the ΛCDM cosmological parameters) times a factor ($f(\sigma)$, see below) that is a function of redshift and overdensity[2] number Δ, but independent of the cosmological parameters. That is the important part because it makes calculating the cluster mass function as a function of cosmology as simple as can be, simple enough to include marginalization over cosmology where cluster counts are needed. That is really all that is needed for the context, but being a bit more precise about what these statements mean, we define the *variance of linear perturbations* as

$$\sigma^2(M,z) = \int P(k)\hat{W}(kR_M)k^2 dk, \tag{1.1}$$

where $P(k)$ is the linear matter power spectrum, and $\hat{W}$ is the Fourier transform of a top-hat window function of mass-related radius

$$R_M = \left(\frac{3M}{4\pi\rho_{mean}(z)}\right)^{1/3}. \tag{1.2}$$

Then the cluster mass function is parameterized in Tinker et al. (2008) as

$$\frac{dn}{dM} = f(\sigma)\frac{\rho_{mean}(z)}{M}\frac{d\ln\sigma^{-1}}{dM}, \tag{1.3}$$

where

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]e^{-c/\sigma^2}. \tag{1.4}$$

[2]Tinker uses the mean density instead of the critical density: $\rho_{mean}(z) = \Omega_M(z)\rho_{crit}(z)$

The value of the parameters A, a, b and c that go into $f(\sigma)$ have been fitted in the simulations, as a function of Δ and z, and Tinker et al. (2008) offers simple fitting functions that reproduce the numerical values of $f(\sigma)$ to an accuracy of 2%.

As explained before, $f(\sigma)$ encodes computationally intensive, complicated but cosmology-independent aspects of the cluster mass function, while the rest of dn/dM is a more tractable function of the ΛCDM cosmological parameters.

Scaling relations

An observable property ζ of clusters of galaxies, for example the velocity dispersion or the SPT SZ significance (see Section 1.2.2), can be related to the cluster mass for a given Δ by a *scaling relation*, of the form

$$\zeta = A \left(\frac{M_\Delta}{M_0} \right)^B f(z), \tag{1.5}$$

where M_0 is a mass pivot, a typical mass for the range where the scaling relation was fit, and $f(z)$ is a redshift-evolution function.

Often, $f(z)$ is related to $H(z)$, the Hubble expansion parameter, therefore

$$\zeta = A \left(\frac{M_{500c}}{M_0} \right)^B \left(\frac{H(z)}{H(z_0)} \right)^C, \tag{1.6}$$

where z_0 is a redshift pivot, and we picked $\Delta = M_{500c}$ for concreteness.

An essential assumption of this parametrization is the the residuals are lognormal. To express this with logarithmic scaling:

$$\ln \zeta = \ln A + B \ln \left(\frac{M_{500c}}{M_0} \right) + C \ln \left(\frac{H(z)}{H(z_0)} \right) + \mathcal{N}(0, D^2) \tag{1.7}$$

where $\mathcal{N}(0, D^2)$ is a normal random variable[3]. The constant D is called the scatter;

[3]this notation is correct if we think of ζ and $500c$ as random variables

sometimes, in loose language, the actual realizations of this normal random variable are called scatter.

The scatter is often quoted in percent, meaning that $D = 0.13$ will be called "13% scatter". It does represent, roughly, a 13% variation in mass:

$$\ln(M/M_0) = \ln(1.0) \pm 0.13 \tag{1.8}$$

$$\Rightarrow M/M_0 = e^{\ln(1.0)\pm 0.13} = e^{\ln(1.0)} e^{\pm 0.13} \tag{1.9}$$

$$= 1.0(1 \pm 0.13 + \frac{0.13^2}{2} + \ldots) \tag{1.10}$$

$$\simeq 1.0(1 \pm 0.13) \tag{1.11}$$

The higher-order terms make this loose but standard language less precise when we are dealing with larger scatter and the higher-order terms cannot be ignored. The mismatch between the two interpretations (i.e. as D, or as a mass percentage) is 0.02 for 20% scatter, and 0.06 for 35% scatter.

Cosmological fit

Generally, the relationship between cluster observable and mass depends on cosmological parameters through the $H(z)$ evolution parameter, and also sometimes through the angular diameter distance $D_A(z)$, which is the conversion factor between observed angular size and physical size, if a physical radius is used for defining the observable, as is the case with the X-ray temperature and Y_X. Therefore the scaling relation has to be fit simultaneously with the cosmological parameters; for cosmology with cluster counts, this is usually done via an MCMC (Vikhlinin et al., 2009b; Mantz et al., 2010; Vanderlinde et al., 2010) . In the case of South Pole Telescope (SPT; see Section 1.2.2) analysis, a modified version of CosmoMC (Lewis & Bridle,

2002) jointly fits CMB data (from WMAP and SPT), cluster counts, and the cluster scaling relations, with priors set from simulations and low-z studies (Benson et al., 2013; Reichardt et al., 2013).

1.2 The hot gas

While most (about 85%) of the mass of any cluster is thought to reside in dark matter, about 90% of the cluster's baryons reside not in galaxies, but in an extended, diffuse halo of gas which is heated up to high temperatures (several to many keV for the most massive clusters) by the dark-matter halo potential. The gas produces X-ray emission via thermal bremsstrahlung, which has been observed for many galaxy clusters by space-based X-ray telescopes. X-ray studies of galaxy clusters is an important and mature field of extragalactic astronomy; see Andersson et al. (2011) and references therein.

1.2.1 The Sunyaev-Zel'dovich effect

In addition to X-ray emission, the hot gas component gives rise to the Sunyaev-Zel'dovich (SZ; Sunyaev & Zel'dovich, 1972) effect. A fraction of the cold photons from the cosmic microwave background (CMB) interact with the hot electrons of the gas (via inverse Compton scattering). As the simple comparison of the energies involved would indicate, after the interaction, the photon has increased its energy. The net effect is that the cluster changes the spectrum of the scattered CMB, where the net number of photons in an energy range is decreased, and the missing photons enhance the signal at higher frequencies. The resulting spectral distortion for a sim-

ulated massive cluster is shown in Figure 1.2; in the case of that cluster, the intensity of the SZ effect signal is of order 5×10^{-4} times the blackbody spectrum component of the CMB.

More precisely, the effect just described is the *thermal* SZ effect. The proper velocity of the cluster with respect to the reference frame of the CMB will also change the CMB spectrum, which is called the *kinectic* SZ effect, but this effect is smaller than the thermal effect and hard to measure. Unlike the thermal SZ effect with has both a decrement and increment of the spectrum, the kinetic SZ effect is indistinguishable from a temperature fluctuaction of the CMB. It has been detected statistically in an ensemble of clusters (Hand et al., 2012).

Unlike the observed luminosity of the X-ray emission of a cluster, which decreases like one over the distance squared, the signal-to-noise ratio of the SZ effect in CMB maps is almost independent of cluster redshift. Indeed, it depends on the angular size of the galaxy cluster, at a given mass. This angular size in turn depends on the angular diameter distance $D_A(z)$, which has units of Megaparsecs per radian; the solid angle of the cluster therefore goes like $1/D_A^2(z)$ which has a relatively weak dependence on redshift above $z \simeq 0.4$ (see Figure 1.3). Thanks to this weak redshift dependence, an SZ survey can therefore yield an essentially volume-complete catalog of clusters (Carlstrom et al., 2002).

CMB surveys such as those conducted by the South Pole Telescope (SPT), the Atacama Cosmology Telescope (ACT), and *Planck* are now reliably finding massive clusters of galaxies through their SZ signature (see, e.g., Staniszewski et al., 2009; Vanderlinde et al., 2010; Williamson et al., 2011; Marriage et al., 2011; Planck Col-

laboration et al., 2011).

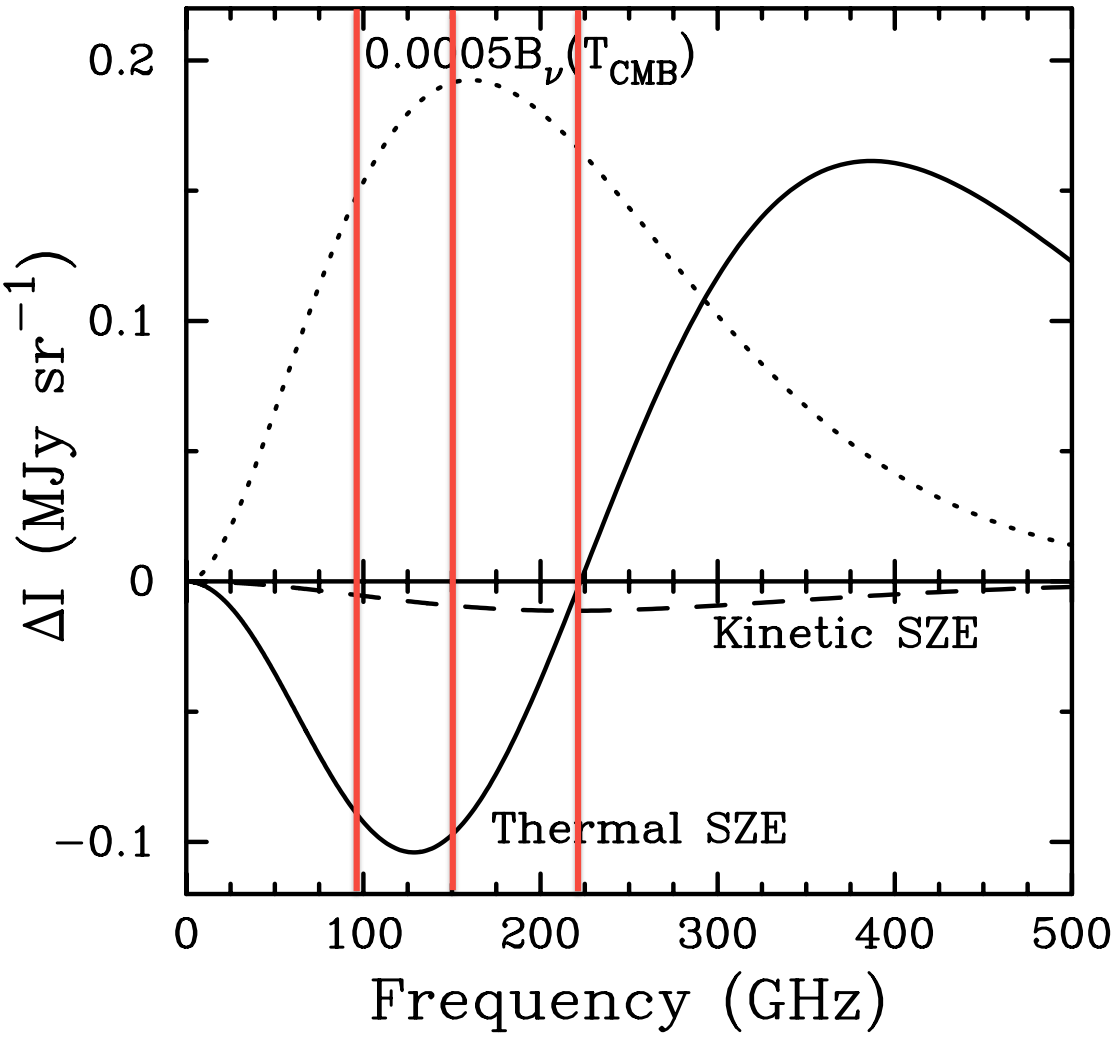

Figure 1.2: The spectral distortion due to the thermal SZ effect for a massive cluster is shown as a solid line. This distortion would be added to the CMB; for reference, the intensity of a black-body spectrum at the temperature of the CMB scaled by a factor of 0.0005 is shown as a dotted line. The dashed line is the kinetic SZ effect distortion if the same cluster has a proper velocity of 500 km s^{-1} The red vertical lines show the approximate centers of the SPT bands for the 2008-2011 observations, at 95, 150 and 220 GHz; the first two are in the SZ decrement, and the last one is at the SZ null.
Image credit: Carlstrom et al. (2002); the red lines were added in the present thesis.

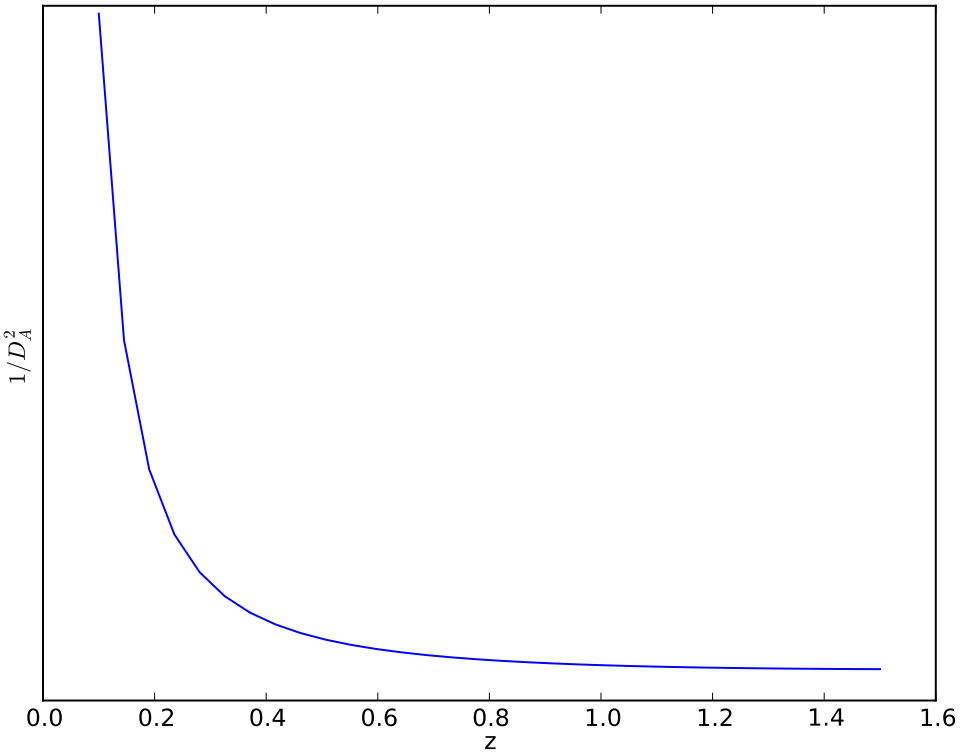

Figure 1.3: Inverse square angular diameter distance $1/D_A^2(z)$ as a function of redshift for a flat ΛCDM cosmology with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.2\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$.

1.2.2 The South Pole Telescope cluster survey

The South Pole Telescope (SPT; Carlstrom et al., 2011) is a 10-meter telescope situated at the South Pole that observes the microwave sky with arcminute resolution. The work presented in this thesis was part of optical follow-up for the SPT-SZ survey, which is a 2500-deg^2 map of the southern sky in 95, 150, and 220 GHz passbands, constructed from SPT observations taken from 2008 to 2011. These passbands correspond respectively to two bands of SZ decrement and one band on the SZ null

(Williamson et al., 2011); this is shown in Figure 1.2. Figure 1.4 shows the footprint of the SPT survey. For a more detailed discussion of the context, and the most recent results from the SPT-SZ cluster survey, see Reichardt et al. (2013) and references therein.

The strength of the SZ signal in the CMB maps from SPT has been characterized by the *detection significance* ξ, the signal-to-noise ratio of the SZ decrement in minimally-filtered maps (see Vanderlinde et al., 2010).

The parameter ξ correlates well with mass, but does not quite follow a scaling relation as described in the previous section, therefore we define the *unbiased significance*

$$\zeta = \sqrt{\xi^2 - 3}. \tag{1.12}$$

The scaling relation of this observable with mass, which we will call the SZ–mass scaling relation, is (Reichardt et al., 2013)

$$\zeta = A_{SZ} \left(\frac{M_{500c}}{3 \times 10^{14} M_{\odot} h^{-1}} \right)^{B_{SZ}} \left(\frac{H(z)}{H(0.6)} \right)^{C_{SZ}}. \tag{1.13}$$

The Reichardt et al. (2013) priors for the cosmological analysis, $A_{SZ} = 6.24$, $B_{SZ} = 1.33$, $C_{SZ} = 0.83$, and $D_{SZ} = 0.24$ for the scatter (with respectively 30%, 20%, 50% and 20% Gaussian uncertainties), are derived from simulations.

We will often refer to the "SZ-mass scaling relation" in later chapters. It refers to this scaling relation, as ζ is the SZ observable that has been used in the SPT cosmological analysis. In theory, a different observable could be used to that end, like the integrated Compton parameter Y_{SZ}, and in that case the Y_{SZ} scaling relation would be "SZ-mass scaling relation".

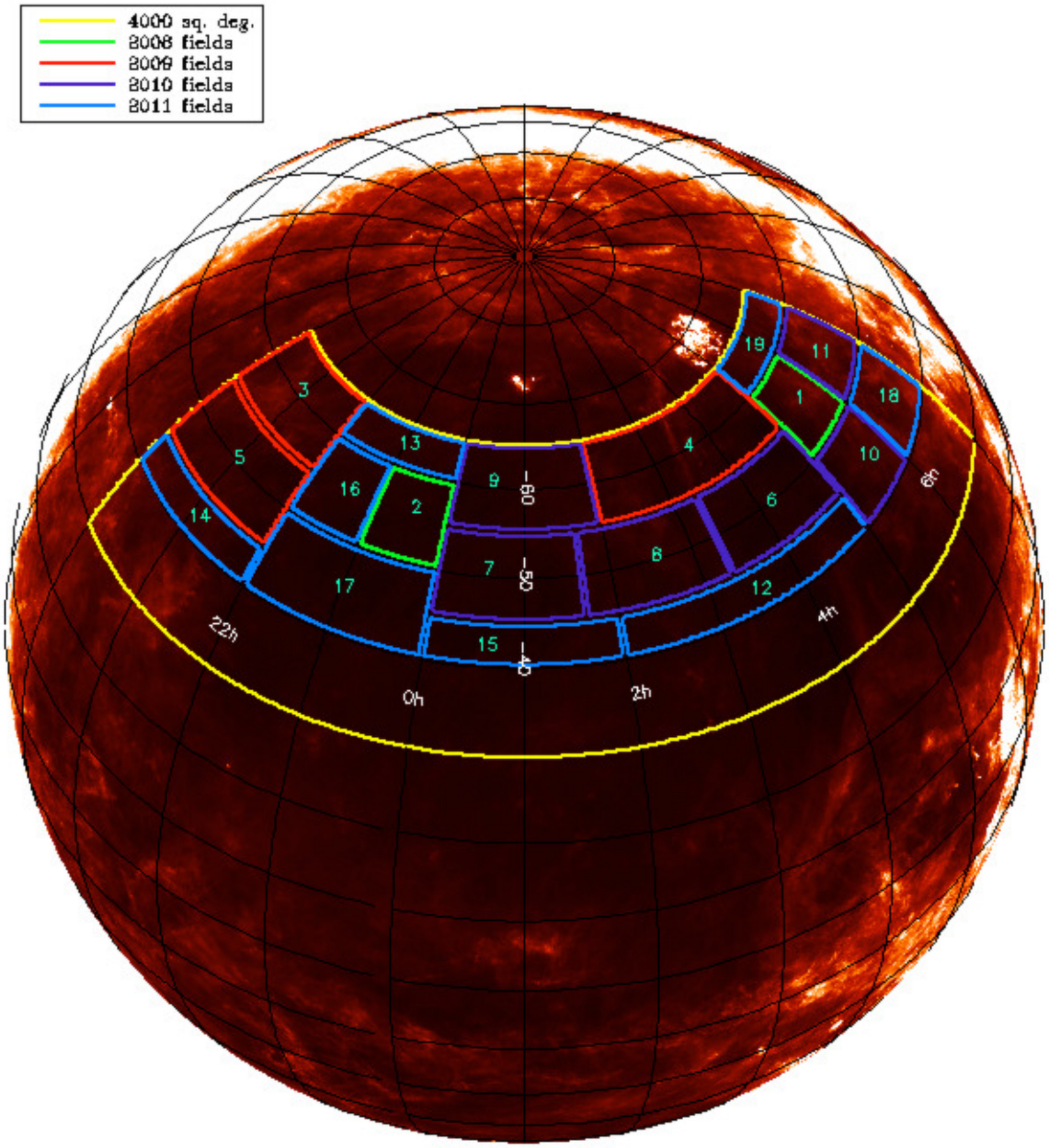

Figure 1.4: SPT survey fields per observation year, overlaid on the IRAS infrared map of the sky.

1.3 Galaxies

The galaxies in galaxy clusters not only show a number-count overdensity with respect to the background density of galaxies, but they also show interesting population features.

There is often a very massive, extended elliptical galaxy that is central to the cluster, called the Brightest Cluster Galaxy (BCG), or just "the central galaxy", depending on whether it is selected purely by magnitude or if morphological criteria are also taken into account (Skibba et al., 2011). The BCG is located close to the bottom of the gravitational potential where it relaxes via dynamical friction, and grows via mergers in the dense cluster center.

In addition to the BCG, many galaxies in the cluster core are red elliptical galaxies, that is galaxies where star formation has mostly stopped; these galaxies form the red sequence of the cluster's color-magnitude diagram (Gilbank et al., 2008). The red sequence forms because the many gravitational interactions in the dense cluster environment shock the gas and trigger star formation in galaxies, so that they become depleted in star-forming gas and evolve to a "red and dead" end-state rapidly when the blue, short-lived, very massive stars die out.

This similarity in color makes the red elliptical members cluster on a color-magnitude diagram. We exploit this in selecting likely cluster members as targets for spectroscopy (see Section 3.3, and Figure 3.4).

Also, this red color is redshifted with distance, and a purely photometric estimate of the cosmological redshift of clusters can be constructed from the measurement of the red-sequence location in the color-magnitude diagram (High et al., 2010; Song

et al., 2012).

1.3.1 Dynamical properties

This thesis deals with measuring the velocity dispersion of clusters found in the SPT survey. A motivation for doing this is presented in the first section of Chapter 2; here we introduce the essential information about velocity dispersions as a cluster mass observable.

The velocity dispersion of a cluster of a given mass can be computed from the first principles of gravity and statistical physics, under certain assumptions about relaxation and equilibrium. The related formulas are the Jeans equation and the virial theorem.

The precision required by contemporary cosmology demands that eventual non-equilibrium effects be taken into account, and importantly, the elongated triaxial nature of dark-matter halos, such that the velocity dispersion to mass scaling relation needs to be calibrated using cosmological N-body simulations rather than found from first principles.

Evrard et al. (2008) has been the *de facto* such calibration. Defining $\sigma_{\rm DM}$ as the velocity dispersion of dark-matter particles within R_{200c}, that is their true, three-dimensional velocity dispersion, not a line-of-sight velocity dispersion, their scaling relation is

$$\sigma_{\rm DM}(M,z) = \sigma_{\rm DM,15}\left[\frac{h(z)M_{200c}}{10^{15}M_\odot}\right]^\alpha \tag{1.14}$$

where

$$\sigma_{\rm DM,15} = 1082.9 \pm 4.0 \ {\rm km\,s^{-1}} \tag{1.15}$$

and

$$\alpha = 0.3361 \pm 0.0026. \tag{1.16}$$

$h(z) = H(z)/100\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ is the normalized Hubble expansion parameter. The scatter in $\ln \sigma_{\rm DM}$ at fixed mass is of order 5%.

Kasun & Evrard (2005) and White et al. (2010) showed that the shape of dark-matter halos introduces a significant dependence of the velocity dispersion on the line of sight. Dark-matter halos in simulations have an elongated, cigar-like triaxial shape, where the two shortest axes are similar and the longest one is about twice as long. This statement is true both in position and velocity space, with the spatial and velocity main axes having a similar orientation, with a typical misalignment of 20-30 degrees. Taking this line-of-sight dependence into account, the scatter in the measurement of the velocity dispersion is nearly 40% in dynamical mass, or $\sim 13\%$ in velocity dispersion at fixed mass.

Saro et al. (2012) calibrated the scaling relation, including the line-of-sight induced scatter, and calculated the velocity dispersion not of dark matter particles, but of dark matter sub halos, identified as galaxies. Their result is, for a scaling relation of the same form as equation 1.14, that

$$\sigma_{\rm DM,15} = 1060 \pm 53\ \mathrm{km\,s^{-1}} \tag{1.17}$$

and

$$\alpha = 0.343 \pm 0.003, \tag{1.18}$$

with the lognormal scatter equal to

$$s_\sigma = \alpha(0.300 + 0.075z). \tag{1.19}$$

This redshift-dependent intrinsic scatter in $\ln \sigma$ at fixed mass is 12% at $z = 0.6$ and reaches 13% above $z = 1$. The part in parentheses in Equation 1.19 is in turn the scatter in the mass at fixed dispersion, and is 35% at the median SPT redshift ($z = 0.62$ in the sample of Reichardt et al., 2013).

1.4 Structure of this thesis

Chapter 2 presents the place of optical spectroscopy in the multi-wavelength follow-up of SPT clusters, and offers mathematical considerations about the effect of survey selection on scaling relations, and simultaneously fitting scaling relations between mass and multiple observables.

Chapter 3 presents considerations surrounding the design of a multi-object spectroscopy program for mass calibration, as well as the detail of the observations that we have completed: targets, instruments used, optical configuration.

Chapter 4 summarizes the procedure for the reduction of the spectroscopic CCD data, and the extraction of galaxy redshifts.

Chapter 5 reviews the different steps necessary to use galaxy redshifts to calculate the cosmological redshift and velocity dispersion of a cluster. We begin with standard methods, and then use resampling to assess the statistical properties of our chosen estimators.

Chapter 6 presents the results of the processing of the data: galaxy redshifts, cluster redshifts, and cluster velocity dispersions. The properties of central galaxies

and a comparison of the data with SZ and X-ray observations are also briefly considered.

Chapter 7 concludes by looking forward to next steps that could be taken to follow this work.

Conventions

Throughout this thesis, we report uncertainties at the 68% confidence level, and we adopt a WMAP7+BAO+H_0 flat ΛCDM cosmology with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.2\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Komatsu et al., 2011). Conversion between M_{500c} and M_{200c} is made assuming an NFW density profile and the Duffy et al. (2008) mass-concentration relation.

Chapter 2

Optical spectroscopy and multi-wavelength follow-up of SPT clusters

2.1 The place of optical spectroscopy

A spectroscopic follow-up program was a necessary component of the SPT cluster survey. Getting the best cosmological parameter constraints from cosmological analyses with clusters require knowledge of each cluster's redshift. While it would be impractical to measure the spectroscopic redshift of all SPT clusters, and the photometric cluster redshifts from the red sequence have small enough uncertainties not to affect the cosmological fit (Vanderlinde et al., 2010), a large spectroscopic subsample provides an essential training set to enable photometric determination of redshifts for the full sample (High et al., 2010; Song et al., 2012).

In practice in the SPT program, the images taken and used for the confirmation of a cluster candidate were often too shallow to yield a reliable photometric redshift at higher redshit ($z \gtrsim 0.9$). Spectroscopic redshifts of the high-redshift end of the sample were therefore important for the cosmology from number counts, especially in the high-mass, high-redshift region of parameter space which is heavily constrained by cosmological models (see Foley et al., 2011). Spectroscopic redshifts at high redshift were also instrumental in planning X-ray (Andersson et al., 2011) and weak lensing (High et al., 2012) follow-up observations, which will yield the best constraints on the calibration of the SZ-mass scaling relation.

As shown in Benson et al. (2013) and Reichardt et al. (2013), the uncertainty in the normalization of the SZ–mass scaling relation (A_{SZ}, in the terminology of these papers and of Section 1.2.2) is the leading source of uncertainty in the cosmological parameter constraints that are derived from SZ-derived cluster abundance measurements. This motivates using multiple mass estimation methods to calibrate the SZ-mass scaling relation, ideally in a joint likelihood analysis. The SPT collaboration is pursuing X-ray observations, weak lensing and velocity dispersions to address the cluster mass calibration challenge. Currently, the relationship between the SZ observable and mass is primarily calibrated in a joint fit of SZ and X-ray data to a model that includes cosmological and scaling relation parameters (Benson et al., 2013, also in Section 1.1). As both the SZ signal and X-ray emission are produced by the hot gas component of the cluster, velocity dispersions and weak lensing are important for assessing any systematic biases from gas-based mass proxies.

How do the eventual results from these methods compare? The intrinsic lognormal

scatter in the mass proxy at fixed true mass is around 18% for $M_{\rm SZ}$ (Reichardt et al., 2013, that is the one which we need to calibrate), 7% for $M_{\rm Y_X}$ (Kravtsov et al., 2006), around 30% for weak-lensing aperture masses (High et al., 2012), and 35% for dynamical masses (Saro et al., 2012). Because the uncertainty on the mean goes as $\sigma/\sqrt{N}$ for N clusters, X-ray calibration would seem to require 20 or 25 times fewer clusters than weak lensing or dynamical masses. However, the absolute mass calibration of the X-ray scaling relation relies on weak lensing mass estimates, so the accuracy is certainly lower than the precision of the scaling relation. Moreover, the scaling relations are calibrated using lower redshift clusters (most of them at $z < 0.1$, see Vikhlinin et al., 2009a) than the typical SPT cluster. The clusters used in the SPT cosmological analysis were at a median redshift of $z = 0.62$ in Reichardt et al. (2013).

Unlike weak-lensing measurements which realistically require space-based imaging at $z \gtrsim 0.6$ and have different observing systematics across the redshift range, velocity dispersions also have the advantage of being obtainable from ground-based telescopes up to high redshift, using similar methods at all redshifts.

2.2 Combining multiple mass observables

This section looks at the scaling relations as random variables and offers a simple mathematical exploration of the implications of selection and comparison of multiple observables. The mathematical presentation is very explicit and therefore may be long in places, but at their heart scaling relations are linear models with a normal random scatter so that the pieces are simple. Section 2.2.1 is a simple description of

the biases due to intrinsic scatter and selection, importantly including the Eddington bias.

Section 2.2.2 describes the notion of Bayesian deboosting. We give it attention because it has proven to be counterintuitive for some colleagues when it comes to follow-up masses.

Section 2.2.3 and the associated Appendix looks at the information that can be extracted from the measured moments (expectation value, variance, covariance) of multiple observables related to the cluster mass. This section is longer and self-contained, and can be omitted on a first reading. It was first explored as a path to reduce the dimensionality of the SPT cosmological MCMC as presented in Vanderlinde et al. (2010), which was not easily scalable to multiple observables. This context is explained at the beginning of the section.

2.2.1 Mass bias from scatter and cluster selection

Fitting the scaling relations requires attention to a couple of statistical biases that arise from the nature of the assumed scaling relation. Assume that we are fitting two different mass estimates against one another, that have independent lognormal scatter about the true mass M:

$$\ln M_1 = \ln M + S_1 \tag{2.1}$$

$$\ln M_2 = \ln M + S_2 \tag{2.2}$$

where S_i is a normal random variable of mean 0 and variance σ_i^2. Then doing an ordinary least squares (OLS) linear regression of $\ln M_1$ vs $\ln M_2$ gives an average

slope of

$$\beta_{\rm OLS} = \frac{\mathrm{Cov}(\ln M_1, \ln M_2)}{\mathrm{Var}(\ln M_2)} = \frac{\mathrm{Cov}(\ln M + S_1, \ln M + S_2)}{\mathrm{Var}(\ln M + S_2)} \tag{2.3}$$

$$= \frac{\mathrm{Var}\ln M}{\mathrm{Var}\ln M + \sigma_2^2} \neq 1. \tag{2.4}$$

The slope is biased away from 1 due to the presence of scatter, and this would be true even when fitting an entire population of clusters, without a selection cut. The selection cut is discussed below. Imposing no selection cut is not a realistic situation but we stress the point that this is not due to selection artifacts.

The second source of bias does arise from the selection process. The clusters are selected as part of a survey, e.g. an SZ survey in the case of SPT. The sample of clusters to be used for cosmology is then defined by a signal-to-noise cut (e.g. $\xi \geq 5.0$ in the case of Vanderlinde et al., 2010), which we will approximate here as being equivalent to a mass cut. The net result of a mass cut M_c, as illustrated in Figure 2.1, is that clusters with $M < M_c$ whose observed mass has scattered up above M_c are kept, while the clusters with $M > M_c$ whose observed mass has scattered down below the cut are rejected. The sample's $\ln M_1$ is biased high compared to $\ln M$, and the slope of linear fits of scaling relations is further affected. This selection effect is the Eddington bias (Eddington, 1913).

These issues are accounted for in a proper statistical treatment, such as the cosmological MCMC fit presented in SPT papers (Benson et al., 2013; Reichardt et al., 2013). The following section tries to clarify some questions related to the Eddington bias when it comes to multi-observable follow-up.

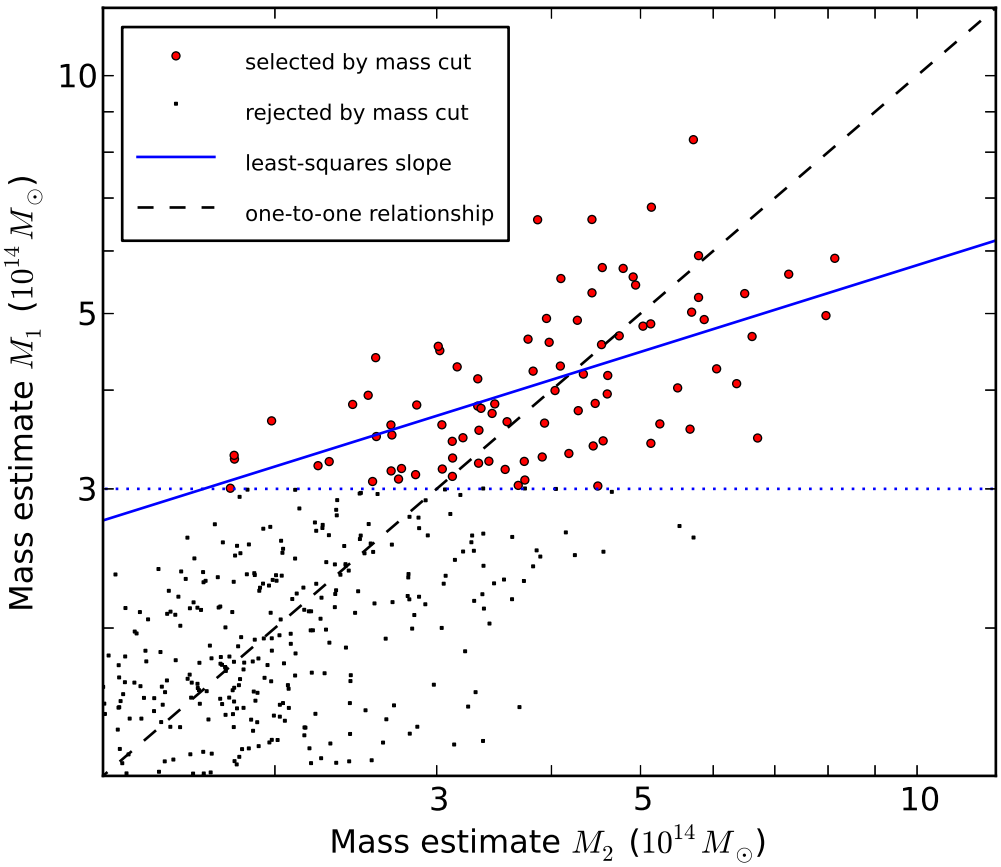

Figure 2.1: Effect of mass selection on mutli-observable fits. "True mass" data points were generated from a toy-model cluster mass function. Mass estimate M_2 has 25% lognormal scatter from the true mass, and mass estimate M_1 has 20% scatter. The mass observables for the entire population are in one-to-one correspondance, but selecting points above a mass cut keeps points that have scattered up above the mass threshold, and rejects points that have scattered down, creating a slope bias.

2.2.2 Follow-up masses for individual clusters

If we measure the velocity dispersion of a cluster that was selected as part of the SPT sample, then does the Eddington bias affect the follow-up measurement? Does the SZ selection offer a prior that needs to be applied to correct the dynamical mass? This last question is relevant in that the SZ observable on which the selection is operated has a smaller scatter than the dynamical mass, so that some of the measured dynamical masses will be outside the range that is "possible" or allowed from the SZ selection. One, a Bayesian perhaps, could then imagine that the dynamical mass can or needs to be corrected in some way.

Part of the answer to all those questions is that one needs to be statistically precise about what question is being asked, and depending on what the question is, properly accounting for the fact that a cluster was SZ-selected in e.g., quoting a dynamical mass from velocity dispersions may or may not involve applying a prior on the measured dynamical masses.

We will think of the follow-up observations in two different ways.

If we have a single interesting cluster, we can ask what is the best estimate of the true mass given the SZ selection and measured velocity dispersion.

If we are instead looking at a sample of follow-up masses, we are interested in the average bias.

Average bias

Suppose that each cluster has a true mass M that is distributed in accord with $\ln M \sim p_M(\ln M)$, where p_M is a properly normalized version of the cluster mass

function, $n_M(M, z)$, described in Section 1.1, also integrated over a range of redshifts for the present argument.

Then there are different observables that yield masses, e.g. the SZ mass and the dynamical mass. They are presumed to relate to the true mass according to the following scaling relations:

$$\ln M_{SZ} = \ln M + S_{SZ} \tag{2.5}$$

$$\ln M_{\rm dyn} = \ln M + S_{\rm dyn} \tag{2.6}$$

where the S are normal random variables whose realisations or standard deviation are usually called *scatter*.

Just to stress what these mean so far for clarity: a realization of those random variables is associated with each cluster, not with each observation. The scatter is due to physical properties and not to measurement errors.

Then we go and observe them. For simplicity, suppose that we select a sample on M_{SZ}, i.e. $M_{SZ} \geq M_0$. From here we need to condition the random variables on this selection.

So, for instance, the expectation value of the mass bias with respect to the true mass in the selected sample is

$$\mathrm{E}(\ln M_{SZ} - \ln M | M_{SZ} \geq M_0) = \mathrm{E}(S_{SZ} | M_{SZ} \geq M_0) \tag{2.7}$$

$$\mathrm{E}(\ln M_{\rm dyn} - \ln M | M_{SZ} \geq M_0) = \mathrm{E}(S_{\rm dyn} | M_{SZ} \geq M_0) \tag{2.8}$$

That is, the expectation value of the scatter, conditioned on the selection.

Let us calculate what these are. The joint probability distribution is

$$p(\ln M = \mu, S_{SZ} = s_{SZ}, S_{\rm dyn} = s_{\rm dyn}) = p_M(\mu) p_N(s_{SZ}, s_{\rm dyn}) \tag{2.9}$$

where p_M is the cluster mass probability function and p_N is a multivariate normal.

The conditional p.d.f. of the scatters once the mass is selected is then

$$\begin{aligned} p(S_{SZ} = s_{SZ},\ S_{\rm dyn} = s_{\rm dyn} | M_{SZ} \geq M_0) & \\ \propto & \int_{M_0}^{\infty} dM_{SZ} \int d\mu \delta(\ln M_{SZ} - \mu - s_{SZ}) p_M(\mu) p_N(s_{SZ}, s_{\rm dyn}) \\ \propto & \int_{M_0}^{\infty} dM_{SZ} p_M(\ln M_{SZ} - s_{SZ}) p_N(s_{SZ}, s_{\rm dyn}) \end{aligned} \tag{2.10}$$

In other words, the mass function is convolved with the multivariate scatter.

In the special case where S_{SZ} and $S_{\rm dyn}$ are uncorrelated, the multivariate normal p.d.f. factors to a product of univariate ones:

$$p_N(s_{SZ}, s_{\rm dyn}) = p_N(s_{SZ}) p_N(s_{\rm dyn}). \tag{2.11}$$

Evaluating the expectation values that we are after, we get

$$\begin{aligned} {\rm E}(S_{SZ} | M_{SZ} \geq M_0) &= \iint ds_{SZ} ds_{\rm dyn} \left[s_{SZ} p(s_{SZ}, s_{\rm dyn} | M_{SZ} \geq M_0) \right] & (2.12) \\ &\propto \int ds_{SZ} \int_{M_0}^{\infty} dM_{SZ} s_{SZ} p_M(\ln M_{SZ} - s_{SZ}) p_N(s_{SZ}) & (2.13) \end{aligned}$$

for the SZ mass Eddington bias, which is nonzero in general. For the other mass:

$$\begin{aligned} {\rm E}(S_{\rm dyn} | M_{SZ} \geq M_0) &= \iint ds_{SZ} ds_{\rm dyn} \left[s_{\rm dyn} p(s_{SZ}, s_{\rm dyn} | M_{SZ} \geq M_0) \right] & (2.14) \\ &\propto \int ds_{\rm dyn} \left[s_{\rm dyn} p_N(s_{\rm dyn}) \right] & (2.15) \\ &= 0. & (2.16) \end{aligned}$$

Therefore, the masses whose scatter is uncorrelated with the selection are unbiased as an ensemble.

Individual cluster masses

There is a different statistical question: what is the expected realization of the scatter given a measured mass? The way that those scatters, adding up to zero in the case of $\mathrm{E}(S_{SZ}|M_{SZ} \geq M_0)$ (see previous section), are distributed across clusters depends on the shape of the selected mass function; to take extreme cases, the clusters with the highest values of M_{dyn} have probably scattered up, while any cluster with, e.g., M_{dyn} lower than the lowest M_{SZ} has probably scattered down. Mathematically, what we want to calculate is

$$\mathrm{E}(S_{\mathrm{dyn}}|M_{SZ} \geq M_0, M_{\mathrm{dyn}}) \tag{2.17}$$

which is different from

$$\mathrm{E}(S_{\mathrm{dyn}}|M_{SZ} \geq M_0), \tag{2.18}$$

which we just calculated.

The p.d.f. that we need to compute this expectation value is (using Bayes' theorem)

$$p(S_{\mathrm{dyn}} = s_{\mathrm{dyn}}|M_{SZ} \geq M_0, \ln M_{\mathrm{dyn}}) = \frac{p(S_{\mathrm{dyn}} = s_{\mathrm{dyn}}, \ln M_{\mathrm{dyn}}|M_{SZ} \geq M_0)}{p(\ln M_{\mathrm{dyn}}|M_{SZ} \geq M_0)} \tag{2.19}$$

Starting again from the simple (2.9):

$$\begin{aligned} p(\ln M_{SZ}, \ln M_{\rm dyn}, S_{\rm dyn} = s_{\rm dyn}) &= \iint ds_{SZ} d\mu \delta(\ln M_{SZ} - \mu - s_{SZ}) \\ &\times \delta(\ln M_{\rm dyn} - \mu - s_{\rm dyn}) \\ &\times p_M(\mu) p_N(s_{SZ}, s_{\rm dyn}) \qquad (2.20) \\ &= p_M(\ln M_{\rm dyn} - s_{\rm dyn}) \\ &\times p_N(\ln M_{SZ} - \ln M_{\rm dyn} + s_{\rm dyn}, s_{\rm dyn}) \qquad (2.21) \end{aligned}$$

Applying the selection:

$$\begin{aligned} p(\ln M_{\rm dyn} = \mu, S_{\rm dyn} = s_{\rm dyn} | M_{SZ} \geq M_0) &= \int_{M_0}^{\infty} dM_{SZ} p_M(\mu - s_{\rm dyn}) \\ &\times p_N(\ln M_{SZ} - \mu + s_{\rm dyn}, s_{\rm dyn}) \qquad (2.22) \end{aligned}$$

Finally, finding the p.d.f. of $\ln M_{\rm dyn}$:

$$\begin{aligned} p(\ln M_{\rm dyn} = \mu | M_{SZ} \geq M_0) &= \int ds_{\rm dyn} \int_{M_0}^{\infty} dM_{SZ} p_M(\mu - s_{\rm dyn}) \\ &\times p_N(\ln M_{SZ} - \mu + s_{\rm dyn}, s_{\rm dyn}) \qquad (2.23) \end{aligned}$$

Therefore, following equation (2.19):

$$\begin{aligned} &p(S_{\rm dyn} = s_{\rm dyn} \,|\, M_{SZ} \geq M_0, \ln M_{\rm dyn} = \mu) \\ &\qquad = \frac{\int_{M_0}^{\infty} dM_{SZ} p_M(\mu - s_{\rm dyn}) p_N(\ln M_{SZ} - \mu + s_{\rm dyn}, s_{\rm dyn})}{\int ds \int_{M_0}^{\infty} dM_{SZ} p_M(\mu - s) p_N(\ln M_{SZ} - \mu + s, s)} \qquad (2.24) \end{aligned}$$

And the expected value of the scatter is

$$\begin{aligned} &\mathrm{E}(S_{\rm dyn} \,|\, M_{SZ} \geq M_0, \ln M_{\rm dyn} = \mu) \\ &\qquad = \frac{\int ds \int_{M_0}^{\infty} dM_{SZ}\, s\, p_M(\mu - s) p_N(\ln M_{SZ} - \mu + s, s)}{\int ds \int_{M_0}^{\infty} dM_{SZ} p_M(\mu - s) p_N(\ln M_{SZ} - \mu + s, s)}. \qquad (2.25) \end{aligned}$$

Then, to check our previous conclusion, the average of the expected scatters will be

$$\begin{aligned}\text{Average } \mathrm{E}\, S_{\rm dyn} &= \sum_{\ln M_{\rm dyn}} \mathrm{E}(S_{\rm dyn}|M_{SZ} \geq M_0, \ln M_{\rm dyn}) P(\ln M_{\rm dyn}|M_{SZ} \geq M_0) \\ &= \int d\mu \int ds \int_{M_0}^{\infty} dM_{SZ}\, s\, p_M(\mu - s) p_N(\ln M_{SZ} - \mu + s, s)\end{aligned} \tag{2.26}$$

Now this can be re-written slightly differently by changing the order of integration, and making the Jacobian transformation $\mu \to \mu + s$:

$$\text{Average } \mathrm{E}\, S_{\rm dyn} = \int_{M_0}^{\infty} dM_{SZ} \iint ds\, d\mu\, s\, p_M(\mu) p_N(\ln M_{SZ} - \mu, s) \tag{2.27}$$

Finally, in the absence of correlation in the scatters, the multivariate normal factorizes into separate normal distributions:

$$\text{Average } \mathrm{E}\, S_{\rm dyn} = \int_{M_0}^{\infty} dM_{SZ} \iint ds\, d\mu\, s\, p_M(\mu) p_N(\ln M_{SZ} - \mu) p_N(s) \tag{2.28}$$

$$= \int_{M_0}^{\infty} dM_{SZ} \int d\mu\, p_M(\mu) p_N(\ln M_{SZ} - \mu) \int ds\, s\, p_N(s) \tag{2.29}$$

$$= 0 \tag{2.30}$$

The removal of $\mathrm{E}(S_{SZ}|M_{SZ})$ is called "Bayesian deboosting", and so we could name the removal of $\mathrm{E}(S_{\rm dyn}|M_{SZ} \geq M_0, \ln M_{\rm dyn})$ "Bayesian deboosting of the follow-up mass". In the case of uncorrelated scatter, this deboosting will not add or subtract ensemble bias because there is none to begin with. Of course it will change the scatter of the sample, as it, so to speak, removes it on average.

2.2.3 Combining multiple mass observables

One can play expectation-value games similar to what was done in the previous section with N generic observables rather than two different masses, which we do in this section, with some of the calculations pushed to an appendix to the chapter.

The goal of this work was to lay the groundwork for a joint fit of scaling relations and cosmology. For $N_{\rm clusters}$ different clusters of galaxies and $N_{\rm observables}$ observables, the dimensionality of the MCMC as presented in Vanderlinde et al. (2010) was scaling as $2^{N_{\rm observables}}$, making it a hard problem to scale, and already impractical to use in its then-current implementation for $N_{\rm observables} = 3$. The idea of the following calculation was to use one observable for cluster counts, and then re-express the observed covariance matrix of the different observables in terms of a small number of parameters describing the effect of the mass function and selection, and the slopes, intercepts and covariances of the scaling relations, effectively making it an $O(N^2_{\rm observables})$ problem.

Scalability of the SPT analysis was achieved in a different way in Benson et al. (2013) and therefore the present idea was never fully implemented with cluster counts, but we include it here because it may be useful, for instance this framework is easily extensible to the simultaneous fit of multiple surveys with different selection. The following lessons or results also emerge from the treatment:

1. As far as the cluster mass function is concerned, a proper fit of the scaling relations only requires knowledge of (and in return, can only constrain) the first and second moments of the "selected mass function", $p_M(\ln M|\xi_i \geq \xi_0)$. Additional constraints on the mass function will come from the number counts fit.

2. $\mathrm{E}(S_\xi|\xi \geq \xi_0) > 0$: the average scatter of the selected data points is positive in the selection observable, in other words, points have scattered up on average.

Summary of the model and results

In our model, we will suppose that for each galaxy cluster, there is a true mass M_i and redshift z_i. In practice we will usually know the cluster redshift. M_i is a realization of the true mass random variable, which is distributed according to the cluster mass function:

$$\ln M, z \sim p_M(\ln M, z). \tag{2.31}$$

There is some freedom as to how to define the true mass and there are different canonical choices for different scaling relations (e.g. M_{200c}, M_{500c}), but we will just assume here that one of them has been chosen for all scaling relations. Of course some mass definition is most natural to each observable, but that difficulty is not limited to the present argument and needs to be addressed in any fit.

The different observables are related to the true mass through scaling relations of the form[1]

$$\theta_i = \ln A_i + \ln f_i(z) + B_i \ln M + S_i \tag{2.32}$$

where i is the index indicating which observable it is, the scatter S_i is normal with variance σ_i^2 and correlations with other scatters ρ_{ij}, and $f_i(z)$ contains both the redshift evolution $E(z)^C$ and other redshift dependence of the observable, for instance luminosity-distance, K-correction, etc.

[1]From here on, observables like θ and ξ will be named in log space to avoid carrying too many logs in the equations.

Now that we have defined the scaling relations, we select our sample from a range of values of an observable that we call ξ,

$$\xi_i \geq \xi_0, \tag{2.33}$$

and call the other observables θ_1, θ_2, etc. We then compute the first and second moments, i.e. the expectation value, variance and covariance of the scaling relations given the selection. Having an expression for these moments in terms of the scaling relation parameters is equivalent in spirit to doing a linear fit.

The explicit calculations can be found in the Appendix to this chapter.

Taylor-expanding the moments to first order in $\rho_{\xi i}$ (more on the motivation later; as we will see, some of the orders are identically zero), we can encode the dependence on the mass function and selection in five constants or parameters; the rest of the parameters in the equations are all from scaling relations. Those five constants are:

$$\widehat{\mathrm{E}}\ln M \equiv \mathrm{E}(\ln M|\xi \geq \xi_0) \tag{2.34}$$

$$\widehat{\mathrm{Var}}\ln M \equiv \mathrm{Var}(\ln M|\xi \geq \xi_0) \tag{2.35}$$

$$\widehat{\mathrm{Cov}}(\ln M, S_\xi) \equiv \mathrm{Cov}(S_\xi, \ln M|\xi \geq \xi_0) \tag{2.36}$$

$$\widehat{\mathrm{E}}S_\xi \equiv \mathrm{E}(S_\xi|\xi \geq \xi_0) \tag{2.37}$$

$$\widehat{\mathrm{Var}}S_\xi \equiv \mathrm{Var}(S_\xi|\xi \geq \xi_0) \tag{2.38}$$

The names given, basically using a hat to remind of the conditionality, aim to make the equations more succinct and remind us that those are constants, while retaining some of the meaning, which names like C_1, C_2 wouldn't do.

We find the basic results that:

$$\mathrm{E}(S_i|\xi \geq \xi_0) = \frac{\rho_{\xi i}\sigma_i}{\sigma_\xi}\widehat{\mathrm{E}}S_\xi + O(\rho_{\xi i}^3) \tag{2.39}$$

$$\mathrm{Cov}(S_i, \ln M|\xi \geq \xi_0) = \frac{\rho_{\xi i}\sigma_i}{\sigma_\xi}\widehat{\mathrm{Cov}}(\ln M, S_\xi) + O(\rho_{\xi i}^3) \tag{2.40}$$

$$\mathrm{Cov}(S_i, S_\xi|\xi \geq \xi_0) = \frac{\rho_{\xi i}\sigma_i}{\sigma_\xi}\widehat{\mathrm{Var}}S_\xi + O(\rho_{\xi i}^3) \tag{2.41}$$

$$\mathrm{Var}(S_i|\xi \geq \xi_0) = \sigma_i^2 + O(\rho_{\xi i}^2) \tag{2.42}$$

$$\mathrm{Cov}(S_i, S_j|\xi \geq \xi_0) = \rho_{ij}\sigma_i\sigma_j + O(\rho^2) \tag{2.43}$$

As will become apparent in the full calculation, the appearance of the first and second moments of the selected mass function does not come from approximating that function as a normal distribution (there is no approximation of the mass function), but rather from the inherently Gaussian nature of scaling relations, that are a linear relation plus a Gaussian scatter.

These basic results can be used to express the full expectation and covariance of arbitrary observables, through the scaling relations.

Appendix: multiple scaling relation statistics

Expectation values and covariances of observables

The previous "basic results", moments of the scatter random variables, cannot be measured directly in the data, therefore in the following, expectation values and covariances of real observables are found. The redshift dependence does not appear as the quantities have been redifined – or corrected for redshift dependence – by $\theta_i \to \theta_i - \ln f_i(z)$. Only in the case of ξ do we write $\tilde{\xi} \equiv \xi - \ln f_\xi(z)$, as reference to the uncorrected ξ still needs to be made, in the selection.

We therefore have, to $O(\rho_{\xi i}^2)$:

$$\begin{aligned} \mathrm{E}(\theta_1|\xi \geq \xi_0) &= \mathrm{E}(\ln A_1 + B_1 \ln M + S_1|\xi \geq \xi_0) && (2.44) \\ &= \ln A_1 + B_1 \widehat{\mathrm{E}} \ln M + \rho_{\xi 1}\sigma_1 \frac{\widehat{\mathrm{E}} S_\xi}{\sigma_\xi} && (2.45) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(\theta_1|\xi \geq \xi_0) &= \mathrm{Var}(\ln A_1 + B_1 \ln M + S_1|\xi \geq \xi_0) && (2.46) \\ &= \mathrm{Var}(B_1 \ln M|\xi \geq \xi_0) + \mathrm{Var}(S_1|\xi \geq \xi_0) + 2\mathrm{Cov}(B_1 \ln M, S_1|\xi \geq \xi_0) \\ &= B_1^2 \widehat{\mathrm{Var}} \ln M + 2B_1 \mathrm{Cov}(\ln M, S_1|\xi \geq \xi_0) + \mathrm{Var}(S_1|\xi \geq \xi_0) \\ &= B_1^2 \widehat{\mathrm{Var}} \ln M + 2B_1 \rho_{\xi 1}\sigma_1 \frac{\widehat{\mathrm{Cov}}(\ln M, S_\xi)}{\sigma_\xi} + \sigma_1^2 && (2.47) \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}(\theta_1, \theta_2|\xi \geq \xi_0) &= \mathrm{Cov}(\ln A_1 + B_1 \ln M + S_1, \ln A_2 + B_2 \ln M + S_2|\xi \geq \xi_0) \\ &= \mathrm{Cov}(B_1 \ln M, B_2 \ln M|\xi \geq \xi_0) + \mathrm{Cov}(B_1 \ln M, S_2|\xi \geq \xi_0) \\ &\quad + \mathrm{Cov}(B_2 \ln M, S_1|\xi \geq \xi_0) + \mathrm{Cov}(S_1, S_2|\xi \geq \xi_0) \\ &= B_1 B_2 \widehat{\mathrm{Var}} \ln M + B_1 \mathrm{Cov}(\ln M, S_2|\xi \geq \xi_0) \\ &\quad + B_2 \mathrm{Cov}(\ln M, S_1|\xi \geq \xi_0) + \mathrm{Cov}(S_1, S_2|\xi \geq \xi_0) \\ &= B_1 B_2 \widehat{\mathrm{Var}} \ln M + (B_1 \rho_{\xi 2}\sigma_2 + B_2 \rho_{\xi 1}\sigma_1) \frac{\widehat{\mathrm{Cov}}(\ln M, S_\xi)}{\sigma_\xi} + \rho_{12}\sigma_1\sigma_2 \\ & && (2.48) \end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(\theta_1, \tilde{\xi}|\xi \geq \xi_0) &= \mathrm{Cov}(\ln A_1 + B_1 \ln M + S_1, \ln A_\xi + B_\xi \ln M + S_\xi|\xi \geq \xi_0) \\
&= B_1 B_\xi \widehat{\mathrm{Var}} \ln M + B_\xi \mathrm{Cov}(\ln M, S_1|\xi \geq \xi_0) \\
&\quad + B_1 \mathrm{Cov}(\ln M, S_\xi|\xi \geq \xi_0) + \mathrm{Cov}(S_1, S_\xi|\xi \geq \xi_0) \\
&= B_1 B_\xi \widehat{\mathrm{Var}} \ln M + \left(B_1 + \frac{B_\xi \rho_{\xi 1} \sigma_1}{\sigma_\xi}\right) \mathrm{Cov}(\ln M, S_\xi|\xi \geq \xi_0) \\
&\quad + \frac{\rho_{\xi 1} \sigma_1}{\sigma_\xi} \mathrm{Var}(S_\xi|\xi \geq \xi_0) \\
&\equiv B_1 B_\xi \widehat{\mathrm{Var}} \ln M + (B_1 \sigma_\xi + B_\xi \rho_{\xi 1} \sigma_1) \frac{\widehat{\mathrm{Cov}}(\ln M, S_\xi)}{\sigma_\xi} \\
&\quad + \rho_{\xi 1} \sigma_1 \frac{\widehat{\mathrm{Var}} S_\xi}{\sigma_\xi} \qquad (2.49)
\end{aligned}$$

We also have exactly, that is not to any order in $\rho_{\xi i}$:

$$\begin{aligned}
\mathrm{E}(\tilde{\xi}|\xi \geq \xi_0) &= \mathrm{E}(\ln A_\xi + B_\xi \ln M + S_\xi|\xi \geq \xi_0) \qquad (2.50) \\
&= \ln A_\xi + B_\xi \widehat{\mathrm{E}} \ln M + \widehat{\mathrm{E}} S_\xi \qquad (2.51)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\tilde{\xi}|\xi \geq \xi_0) &= \mathrm{Var}(\ln A_\xi + B_\xi \ln M + S_\xi|\xi \geq \xi_0) \qquad (2.52) \\
&= B_\xi^2 \widehat{\mathrm{Var}} \ln M + \mathrm{Var}(S_\xi|\xi \geq \xi_0) + 2 B_\xi \mathrm{Cov}(\ln M, S_\xi|\xi \geq \xi_0) \\
&\equiv B_\xi^2 \widehat{\mathrm{Var}} \ln M + \widehat{\mathrm{Var}} S_\xi + 2 B_\xi \widehat{\mathrm{Cov}}(\ln M, S_\xi) \qquad (2.53)
\end{aligned}$$

Calculations: the easy way

We start with a multivariate normal:

$$p_N(S = s, S_\xi = s_\xi) = \frac{1}{2\pi\sigma\sigma_\xi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{s^2}{\sigma^2} - \frac{2\rho s s_\xi}{\sigma\sigma_\xi} + \frac{s_\xi^2}{\sigma_\xi^2}\right]\right) \qquad (2.54)$$

To linear order in ρ:

$$p_N(s, s_\xi) \simeq p_N(s)p_N(s_\xi)\exp\left(\frac{\rho s s_\xi}{\sigma\sigma_\xi}\right) \quad (2.55)$$

$$\simeq p_N(s)p_N(s_\xi)\left(1+\frac{\rho s s_\xi}{\sigma\sigma_\xi}\right) \quad (2.56)$$

We can substitute that expansion in the joint p.d.f.:

$$\begin{aligned} &p(\ln M = \mu, S = s, S_\xi = s_\xi \,|\, \xi \geq \xi_0) \\ &\quad = \iint_{\xi_0}^{\infty} p(\xi|\mu, z, s_\xi)p_N(s_\xi, s)p_M(\mu, z)dzd\xi \quad (2.57) \\ &\quad = \iint_{\xi_0}^{\infty} p(\xi|\mu, z, s_\xi)p_N(s)p_N(s_\xi) \\ &\quad \times \left(1+\frac{\rho s s_\xi}{\sigma\sigma_\xi}\right) p_M(\mu, z)dzd\xi \end{aligned}$$

where, by slight abuse of notation,

$$p(\xi|\mu, z, s_\xi) \sim \delta(\xi - \ln f_\xi(z) - B_\xi\mu - s_\xi). \quad (2.58)$$

"Slight abuse of notation" means here in practice that we will need to be cautious about the order of integration, so that either s_ξ or μ (or both) need to be integrateed over, and need to be integrated before ξ. Choosing to integrate over μ, we get

$$\begin{aligned} p(S = s, S_\xi = s_\xi|\xi \geq \xi_0) &= p_N(s)\left(1+\frac{\rho s s_\xi}{\sigma\sigma_\xi}\right)p_N(s_\xi) \\ &\times \iint_{\xi_0}^{\infty} p_M\left(\frac{\xi - \ln f_\xi(z) - s_\xi}{B_\xi}, z\right)dzd\xi \quad (2.59) \end{aligned}$$

Integrating over s, we get

$$p(S_\xi = s_\xi|\xi \geq \xi_0) = p_N(s_\xi)\iint_{\xi_0}^{\infty} p_M\left(\frac{\xi - \ln f_\xi(z) - s_\xi}{B_\xi}, z\right)dzd\xi \quad (2.60)$$

therefore,

$$p(S = s, S_\xi = s_\xi|\xi \geq \xi_0) = p_N(s)\left(1+\frac{\rho s s_\xi}{\sigma\sigma_\xi}\right)p(s_\xi|\xi \geq \xi_0) \quad (2.61)$$

We can now calculate:

$$\begin{aligned} \mathrm{E}(S|\xi \geq \xi_0) &= \iint s p_N(s) \left(1 + \frac{\rho s s_\xi}{\sigma \sigma_\xi}\right) p(s_\xi|\xi \geq \xi_0) ds ds_\xi & (2.62) \\ &= \frac{\rho\sigma}{\sigma_\xi} \mathrm{E}(S_\xi|\xi \geq \xi_0) & (2.63) \end{aligned}$$

In the next subsection, where we do the integrals more explicitly, we will calculate that this expectation value is $\rho\sigma C$ with $C \geq 0$. With this condition translates here to $\mathrm{E}(S_\xi|\xi \geq \xi_0) \geq 0$, which makes sense intuitively: because of the cut $\xi \geq \xi_0$, some points that have scattered up across the cut will be kept, and some points that have scattered down across the cut will not be selected, leaving an overall positive expectation value for the scatter.

We can also calculate the covariance here:

$$\begin{aligned} \mathrm{E}(S_\xi S|\xi \geq \xi_0) &= \iint s s_\xi p_N(s) \left(1 + \frac{\rho s s_\xi}{\sigma \sigma_\xi}\right) p(s_\xi|\xi \geq \xi_0) & (2.64) \\ &= \rho\sigma \cdot \frac{1}{\sigma_\xi} \mathrm{E}(S_\xi^2|\xi \geq \xi_0) & (2.65) \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{Cov}(S_\xi, S|\xi \geq \xi_0) &= \mathrm{E}(S_\xi S|\xi \geq \xi_0) - \mathrm{E}(S|\xi \geq \xi_0)\mathrm{E}(S_\xi|\xi \geq \xi_0) & (2.66) \\ &= \rho\sigma \cdot \frac{1}{\sigma_\xi} \left(\mathrm{E}(S_\xi^2|\xi \geq \xi_0) - \mathrm{E}(S_\xi|\xi \geq \xi_0)^2\right) & (2.67) \\ &= \frac{\rho\sigma}{\sigma_\xi} \mathrm{Var}(S_\xi|\xi \geq \xi_0). & (2.68) \end{aligned}$$

Starting again with equation (2.57), but this time integrating over s_ξ, we have

$$\begin{aligned} p(\mu, s|\xi \geq \xi_0) &= \iint_{\xi_0}^{\infty} p_N(s) p_N(\xi - \ln f_\xi(z) - B_\xi \mu) \\ &\times \left(1 + \frac{\rho s (\xi - \ln f_\xi(z) - B_\xi \mu)}{\sigma \sigma_\xi}\right) p_M(\mu, z) dz d\xi & (2.69) \end{aligned}$$

The piece that will be needed to interpret the coming result is starting from

$$p(\ln M = \mu, S_\xi = s_\xi | \xi \geq \xi_0) = \iint_{\xi_0}^{\infty} p(\xi|\mu, z, s_\xi) p_N(s_\xi) p_M(\mu, z) dz d\xi \tag{2.70}$$

and being careful about the order of integration:

$$\begin{aligned} \mathrm{E}(S_\xi \ln M | \xi \geq \xi_0) &= \iiiint_{\xi_0}^{\infty} \mu\, s_\xi\, p(\xi|\mu, z, s_\xi) p_N(s_\xi) p_M(\mu, z) dz d\xi d\mu ds_\xi \quad (2.71) \\ &= \iiint_{\xi_0}^{\infty} \mu\, (\xi - \ln f_\xi(z) - B_\xi \mu)\, p_N(\xi - \ln f_\xi(z) - B_\xi \mu) \\ &\times \; p_M(\mu, z) dz d\xi d\mu \end{aligned}$$

From (2.69), we can calculate

$$\begin{aligned} \mathrm{E}(S \ln M | \xi \geq \xi_0) &= \iint s\mu \iint_{\xi_0}^{\infty} p_N(s) p_N(\xi - \ln f_\xi(z) - B_\xi \mu) \\ &\times \; \left(1 + \frac{\rho s (\xi - \ln f_\xi(z) - B_\xi \mu)}{\sigma \sigma_\xi}\right) p_M(\mu, z) dz d\xi \\ &= \frac{\rho\sigma}{\sigma_\xi} \iiint_{\xi_0}^{\infty} \mu f(\xi) p_N(\xi - \ln f_\xi(z) - B_\xi \mu) \\ &\times \; (\xi - \ln f_\xi(z) - B_\xi \mu) p_M(\mu, z) dz d\xi d\mu \\ &= \frac{\rho\sigma}{\sigma_\xi} \mathrm{E}(S_\xi \ln M | \xi \geq \xi_0) \quad (2.72) \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{Cov}(S, \ln M | \xi \geq \xi_0) &= \frac{\rho\sigma}{\sigma_\xi} \mathrm{E}(S_\xi \ln M | \xi \geq \xi_0) \quad (2.73) \\ &\quad - \mathrm{E}(\ln M | \xi \geq \xi_0) \cdot \frac{\rho\sigma}{\sigma_\xi} \mathrm{E}(S_\xi | \xi \geq \xi_0) \quad (2.74) \\ &= \frac{\rho\sigma}{\sigma_\xi} \mathrm{Cov}(S_\xi, \ln M | \xi \geq \xi_0) \quad (2.75) \end{aligned}$$

Explicit integrals

To show that the Taylor expanson of the previous section is a desirable thing to do, we start here by going as far as possible with an exact calculation.

Just like what was done in Section 2.2.2, we will start with the joint probability distribution:

$$\begin{aligned} p(\ln M = \mu, S = s, \xi = \xi) &= \iint \delta(\xi - B_\xi \mu - \ln f_\xi(z) - s_\xi) p_N(s_\xi, s) p_M(\mu, z) dz ds_\xi \\ &= \int p_N(\xi - B_\xi \mu - \ln f_\xi(z), s) p_M(\mu, z) dz \end{aligned} \tag{2.76}$$

where S is the scatter of some observable, p_N is a multivariate normal with variances σ_ξ, σ and correlation ρ, and p_M is the cluster mass function; also, for concision, $\ln A_\xi$ has been absorbed into $\ln f_\xi(z)$.

From there, we need to integrate to find the conditional distribution:

$$p(\ln M = \mu, S = s | \xi \geq \xi_0) = \iint_{\xi_0}^{\infty} p_N(\xi - B_\xi \mu - \ln f_\xi(z), s) p_M(\mu, z) dz d\xi \tag{2.77}$$

Here we need to be a little more explicit. The bivariate normal distribution, the joint p.d.f. of scatters, is

$$p_N(s_\xi, s) = \frac{1}{2\pi\sigma\sigma_\xi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{s^2}{\sigma^2} - \frac{2\rho s s_\xi}{\sigma\sigma_\xi} + \frac{s_\xi^2}{\sigma_\xi^2}\right]\right) \tag{2.78}$$

Therefore, writing the constant at the front of p_N as K, we have the integral

$$\begin{aligned}
&\int_{\xi_0}^{\infty} p_N(\xi - B_\xi\mu - \ln f_\xi(z), s)d\xi \\
&\quad = K\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\frac{s^2}{\sigma^2}\right) \\
&\quad \times \int_{\xi_0}^{\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left[-\frac{2\rho s\left(\xi - B_\xi\mu - \ln f_\xi(z)\right)}{\sigma\sigma_\xi}\right.\right. \\
&\qquad \left.\left. + \frac{\left(\xi - B_\xi\mu - \ln f_\xi(z)\right)^2}{\sigma_\xi^2}\right]\right)d\xi \\
&\quad = K\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\frac{s^2}{\sigma^2}\right) \\
&\quad \times \int_{\xi_0 - B_\xi\mu - \ln f_\xi(z)}^{\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left[-\frac{2\rho s\xi}{\sigma\sigma_\xi} + \frac{\xi^2}{\sigma_\xi^2}\right]\right)d\xi \\
&\quad = K\exp\left(-\frac{1-\rho^2}{2\left(1-\rho^2\right)}\frac{s^2}{\sigma^2}\right) \\
&\quad \times \int_{\xi_0 - B_\xi\mu - \ln f_\xi(z)}^{\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{\xi}{\sigma_\xi} - \frac{\rho s}{\sigma}\right]^2\right)d\xi \\
&\quad = K\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \\
&\quad \times \sigma_\xi\int_{\frac{\xi_0 - B_\xi\mu - \ln f_\xi(z)}{\sigma_\xi} - \frac{\rho s}{\sigma}}^{\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\hat{\xi}^2\right)d\hat{\xi} \\
&\quad = K\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \\
&\quad \times \sigma_\xi\frac{\sqrt{\pi}\sqrt{2(1-\rho^2)}}{2}\mathrm{erfc}\left[\frac{1}{\sqrt{2(1-\rho^2)}}\left(\frac{\xi_0 - B_\xi\mu - \ln f_\xi(z)}{\sigma_\xi} - \frac{\rho s}{\sigma}\right)\right]
\end{aligned}$$

In summary,

$$\begin{aligned}
\int_{\xi_0}^{\infty} p_N(\xi - B_\xi\mu - \ln f_\xi(z), s)d\xi &= \frac{1}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \\
&\times \mathrm{erfc}\left[\frac{1}{\sqrt{2(1-\rho^2)}}\left(\frac{\xi_0 - B_\xi\mu - \ln f_\xi(z)}{\sigma_\xi} - \frac{\rho s}{\sigma}\right)\right]
\end{aligned} \tag{2.79}$$

and

$$\begin{aligned} p(\ln M = \mu, S = s|\xi \geq \xi_0) &= \frac{N}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \\ &\times \int \mathrm{erfc}\left[\frac{1}{\sqrt{2(1-\rho^2)}}\left(\frac{\xi_0 - B_\xi\mu - \ln f_\xi(z)}{\sigma_\xi} - \frac{\rho s}{\sigma}\right)\right] \\ &\times p_M(\mu, z)dz \end{aligned} \tag{2.80}$$

This is as much as we can do exactly, and that remaining integral could be done numerically, but we can continue from here if we do a Taylor expansion in ρ. It is a reasonable expansion to do; among other reasons (but that will no longer be as true for more than two observables), if the scatters of our only two observables are very correlated, then we cannot hope to untangle all of their parameters, as a correlated realization of the scatter just looks like a different mass.

The complementary error function is linked to the error function by

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) \tag{2.81}$$

The derivative of the error function is

$$\frac{\mathrm{d}\,\mathrm{erf}(x)}{\mathrm{d}x} = \frac{2}{\sqrt{\pi}} \exp(-x^2) \tag{2.82}$$

therefore

$$\mathrm{erf}(a+\epsilon) = \mathrm{erf}(a) + \left.\frac{\mathrm{d}\,\mathrm{erf}(x)}{\mathrm{d}x}\right|_{x=a} \epsilon + \ldots \tag{2.83}$$

$$\simeq \mathrm{erf}(a) + \frac{2}{\sqrt{\pi}} \exp(-a^2)\epsilon \tag{2.84}$$

So

$$\begin{aligned}
& \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\left(\frac{\xi_0 - \ln f_\xi(z) - B_\xi\mu}{\sigma_\xi} - \frac{\rho s}{\sigma}\right)\right] \\
= \;& 1 - \operatorname{erf}\left(\frac{\xi_0 - \ln f_\xi(z) - B_\xi\mu}{\sqrt{2}\sigma_\xi} - \frac{\rho s}{\sqrt{2}\sigma}\right) \\
\simeq \;& 1 - \operatorname{erf}\left(\frac{\xi_0 - \ln f_\xi(z) - B_\xi\mu}{\sqrt{2}\sigma_\xi}\right) + \frac{2}{\sqrt{\pi}}\exp\left[-\left(\frac{\xi_0 - \ln f_\xi(z) - B_\xi\mu}{\sqrt{2}\sigma_\xi}\right)^2\right]\frac{\rho s}{\sqrt{2}\sigma}
\end{aligned} \tag{2.85}$$

We therefore write, to leading order in ρ:

$$\begin{aligned}
p(\ln M = \mu, S = s|\xi \geq \xi_0) \;&=\; \int \left(C_0(\mu,z) + C_1(\mu,z)\frac{\rho s}{\sigma}\right) \\
&\times\; \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) p_M(\mu,z)dz
\end{aligned} \tag{2.86}$$

Notice that $C_1(\mu, z) > 0$ as it is an exponential.

Also, integrating over s, we find

$$p(\ln M = \mu|\xi \geq \xi_0) = \int C_0(\mu,z)p_M(\mu,z)dz \tag{2.87}$$

$C_0 \times p_M$ therefore represents the selected mass function. Because the total probability is 1,

$$\iint C_0(\mu,z)p_M(\mu,z)d\mu dz = 1 \tag{2.88}$$

Integrating first over μ instead, we find the conditional distribution of S:

$$p(S = s|\xi \geq \xi_0) = \left(1 + C\rho\frac{s}{\sigma}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \tag{2.89}$$

where $C > 0$. We saw previously that this $C \propto \mathrm{E}(S_\xi|\xi \geq \xi_0)$, so that the condition $C > 0$ has a natural interpretation.

Expectation values and covariances

We can now compute the conditional moments of our random variables.

$$\mathrm{E}(S|\xi \geq \xi_0) = \int s\, p(S=s|\xi \geq \xi_0) ds = C\rho \frac{\mathrm{Var}S}{\sigma} + O(\rho^3) \tag{2.90}$$

$$= C\rho \frac{\sigma^2}{\sigma} + O(\rho^3) \tag{2.91}$$

$$= C\rho\sigma + O(\rho^3) \tag{2.92}$$

The ρ^2 term is identically 0 as our unconditioned normal distribution has mean 0 and doesn't have a third moment. Then

$$\mathrm{Var}(S|\xi \geq \xi_0) = \mathrm{E}(S^2|\xi \geq \xi_0) - (\mathrm{E}(S|\xi \geq \xi_0))^2 \tag{2.93}$$

$$\mathrm{E}(S^2|\xi \geq \xi_0) = \int s^2\, p(S=s|\xi \geq \xi_0) ds \tag{2.94}$$

$$= \mathrm{Var}S + O(\rho^2)$$

$$= \sigma^2 + O(\rho^2) \tag{2.95}$$

$$\mathrm{Var}(S|\xi \geq \xi_0) = \sigma^2 + O(\rho^2) - (C\rho\sigma)^2$$

$$= \sigma^2 + O(\rho^2) \tag{2.96}$$

Also,

$$\mathrm{E}(\ln M\, S|\xi \geq \xi_0) = \iint \mu\, s\, p(\mu, s|\xi \geq \xi_0) \tag{2.97}$$

$$= \frac{1}{\sqrt{2}}\rho\sigma \iint \mu C_1(\mu) p_M(\mu, z) d\mu dz \tag{2.98}$$

So that

$$\begin{aligned}\mathrm{Cov}(\ln M, S|\xi \geq \xi_0) &= \frac{1}{\sqrt{2}}\rho\sigma \iint \mu C_1(\mu, z) p_M(\mu, z) d\mu dz \\ &\quad -\frac{-C_1}{\sqrt{2}}\rho\sigma \iint \mu C_0(\mu, z) p_M(\mu, z) d\mu dz \\ &\propto \rho\sigma. \end{aligned} \tag{2.99}$$

Finally, doing the above calculation with two observables in addition to ξ (which involves a three-dimensional normal distribution), we find that

$$\mathrm{Cov}(S_1, S_2|\xi \geq \xi_0) = \rho_{12}\sigma_1\sigma_2 + O(\rho^2). \tag{2.100}$$

Chapter 3

A spectroscopic follow-up program for the SPT-SZ cluster survey

This chapter details both the strategy of the SPT cluster spectroscopic follow-up, in Section 3.1, and the specifics of the observations we have carried out, in Sections 3.2 and 3.3.

3.1 A few-$N_{\rm members}$ spectroscopic strategy

The design of a spectroscopic follow-up program for clusters hinges on the fact that contemporary multi-object spectrographs use slit masks, so that the investment in telescope time, the limiting resource for the follow-up, is quantized by how many masks are allocated to each cluster. The optimization problem is therefore to allocate the observation of M masks (a fixed number) across C clusters. We want to do it in a way that minimizes the uncertainty in the ensemble cluster mass normalization,

from dynamical masses. Considerations surrounding the measurement of the cluster's redshift, the other important goal of the follow-up program, do factor in, as will be discussed below, but do not affect the optimization discussed here since a reliable redshift can be obtained with a single mask.

We use a simple model to carry out this optimization. We call the number of member galaxy redshifts obtained in a single cluster $N_{\rm members}$; it is an important assumption of our model that these $N_{\rm members}$ have been identified as members of the galaxy cluster. We will revisit this assumption below. We call the average number of member redshifts observed per multi-slit mask r ("return"), and the number of masks per clusters m. Therefore, on average we have that

$$N_{\rm members} = rm \tag{3.1}$$

and

$$M = Cm. \tag{3.2}$$

Another simplifying assumption that we make is that the distribution of velocities in a cluster is close to a normal distribution. That is certainly true on average, as we will see with the stacked cluster in Section 5.2, but the measured distribution for individual clusters can deviate from normality. The use of robust estimators minimizes the impact of such deviations on the measured redshift and dispersion. We will use the fact that the variance of the mean and variance (dispersion squared) for a Gaussian probability distribution are

$$\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{C}, \tag{3.3}$$

$$\mathrm{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{C-1}, \tag{3.4}$$

where the caret denotes the statistical estimator. Per Equation 3.4, the fractional uncertainty in the velocity dispersion of a single cluster will be

$$\sigma_{\rm dispersion} = \frac{1}{2}\sqrt{\frac{2}{N_{\rm members}-1}}, \tag{3.5}$$

where potentially confusing notation has been used to keep up with existing conventions: σ in equations 3.3 and 3.4 is the standard deviation (dispersion) of the normal distribution (the distribution of velocities) while $\sigma_{\rm quantity}$ is the fractional uncertainty in a quantity. Since the dynamical mass is proportional to the dispersion cubed, the fractional uncertainty in the dynamical mass is three times $\sigma_{\rm dispersion}$. N-body simulations inform us that line-of-sight projection effects induce an unavoidable scatter in the relationship between line-of-sight velocity dispersion and cluster mass (Kasun & Evrard, 2005; White et al., 2010); this intrinsic scatter is 12% in ln(dispersion) at fixed mass, implying a 35% scatter in dynamical mass (Saro et al., 2012). This 35% intrinsic scatter needs to be added in quadrature to the dynamical mass uncertainty of any one cluster; as we will see, it ultimately implies that obtaining higher-precision velocity dispersions on a few clusters is less informative (for the purposes of mass calibration) than obtaining coarser dispersions on more clusters. The fractional uncertainty on the mass of a single cluster is given by

$$(\sigma_M)^2 = \left(\frac{3}{2}\sqrt{\frac{2}{N_{\rm members}-1}}\right)^2 + (0.35)^2\,. \tag{3.6}$$

The fractional uncertainty in the mass normalization, the mean of C masses, will therefore be

$$(\sigma_{\rm norm})^2 = \frac{(\sigma_M)^2}{C} = \frac{1}{C}\left[\left(\frac{3}{2}\sqrt{\frac{2}{rm-1}}\right)^2 + (0.35)^2\right] \tag{3.7}$$

where we have replaced $N_{\rm members} = rm$. It may already be clear to the reader that C simply needs to be as large as possible. But m also varies in a way that is tied to C. As stated in Equation 3.2, we know that the total number of masks $M = Cm$, therefore $C = M/m$ and

$$(\sigma_{\rm norm})^2(m) = \frac{m}{M}\left[\left(\frac{3}{2}\sqrt{\frac{2}{rm-1}}\right)^2 + (0.35)^2\right]. \tag{3.8}$$

This function is minimized for $m = 7.1/r$. The average number of members per mask, r, is certainly greater than 7 for the spectrographs that we have been using, approximately in the $10-15$ range up to high redshift (and would be greater if we considered it separately in the low to medium redshift range, $z \lesssim 0.8$.) Therefore, $m \lesssim 1$.

So from this argument alone, the right thing to do is to is to observe a single mask per cluster. However, we should remember the assumptions that we made. The first assumption was that the measured galaxies were cluster members. The membership determination is done from the observed distribution of velocities, therefore its efficiency will depend on the number of members in the observed sample of galaxies. In other words, the observation needs to be good enough for a proper discrimination of the main distribution from the background. The experience encapsulated in the velocity dispersion literature is that 7 galaxies is too few for determining a reliable velocity dispersion, and a cutoff of around 20 (e.g. Girardi et al., 1993), 25, 30 or more (e.g. Girardi et al., 1996; Zhang et al., 2011), is usually used in studies of ensembles.

We can rephrase the lesson learned from the above calculation as follows: we need to observe as few masks per cluster as is necessary for a reliable determination of cluster membership. Considering this result and following the cited velocity dispersion

literature, we have adopted a target of $N_{\rm members} \sim 20-30$ to yield a reliable dispersion for a cluster. To achieve this target range of $N_{\rm members}$, we generally need to observe two masks per cluster on the spectrographs available to us. The use of a red-sequence color selection is a necessary feature of this strategy to target likely cluster members and maximize the cluster members per mask (r, as defined above), as a small number of multislit masks only allows us to target a small fraction of the galaxies in the region of the sky around the SZ center.

As a side note, from the discussion above, in particular equation 3.6, we see that there is no important statistical gain in the knowledge of the mass of a single cluster when $N_{\rm members}$ is larger than the value for which

$$\frac{3}{2}\sqrt{\frac{2}{N_{\rm members}-1}} = 0.35 \tag{3.9}$$

$$\Rightarrow N_{\rm members} = 38. \tag{3.10}$$

Of course data on more members could be useful for studies other than the mass, like substructure, the study of mass systematics, or galaxy evolution.

Note that while obtaining two masks per cluster is our strategy in view of mass calibration, some of the observations presented in Section 3.3 depart from that model and have only one mask observed, with correspondingly fewer members. In some cases, the second mask has yet to be observed, and other observations come from a number of programs with different objectives, for instance the identification and characterization of high-redshift clusters, the follow-up of bright sub-millimeter galaxies, and long slit observations from the early days of our follow-up program. Finally, some clusters of special interest were targeted with more than 2 masks. The spectroscopic follow-up of high-redshift clusters to obtain their redshift has been very important

for many aspects of the SPT multi-wavelength follow-up. The limited telescope time available for the optical imaging of SPT clusters meant that the observations were often just deep enough for confirmation (or, the confirmation came from *Spitzer*/IRAC infrared imaging) but too shallow for an accurate red-sequence photometric redshift, for clusters at $z \gtrsim 0.8$, and certainly at $z > 1.0$, spectroscopy was not an inefficient way to get a reliable redshift.

A note about $N_{\rm members} \geq 15$: In discussions throughout this document, especially in Chapter 6, we often use a $N_{\rm members} \geq 15$ cut for "reliable" dispersions. We note that this number is chosen somewhat arbitrarily, in line with literature in the field, for the conservative exclusion of systems with very few members. As we will see, especially in the resampling analysis of section 5.3, there is no overwhelming statistical evidence for using that exact number, except that for fewer members, the errors and the evaluation of confidence intervals are not as well-behaved.

3.2 Observations: South Pole Telescope

Most of the galaxy clusters for which we report spectroscopic observations (our own measurements for 60 of them and 20 from the literature, with 5 overlapping) were published as SPT cluster detections (and new discoveries) in Vanderlinde et al. (2010), Williamson et al. (2011), and Reichardt et al. (2013); we refer the reader to those publications for details of the SPT observations. The SPT IDs of the clusters and their essential SZ properties are presented in Table 3.1. They include the right ascension and declination of the SZ center, the cluster redshift (from optical spectroscopy, included here for reference), and the SPT detection significance ξ. For those

clusters at redshift $z \geq 0.3$, which is the range used for the SPT cosmological analysis (Vanderlinde et al., 2010), we also report $M_{500c,\mathrm{SZ}}$, the mass estimate from the SPT analysis (Reichardt et al., 2013). It is important to keep in mind that $M_{500c,\mathrm{SZ}}$ is determined from either the SZ data alone or the joint mass estimate from the SZ plus X-ray data, where X-ray measurements are available; the perhaps misleading name is chosen to be in line with SPT publications. The uncertainty in $M_{500c,\mathrm{SZ}}$ includes uncertainty in the scaling relation and cosmological parameters, as described by Benson et al. (2013).

There are 12 clusters that do not appear in prior SPT publications, and are presented here as SPT detections for the first time. Five of them are new discoveries (identified with * in Table 3.1), and the other seven were previously published as ACT detections (Marriage et al., 2011, identified with ** in Table 3.1). The associated SPT observations will be reported in a future SPT cluster catalog paper for the complete 2500-deg^2 SPT-SZ survey, as part of a much larger set of new SPT detections and cluster discoveries.

Table 3.1: SPT properties and source of spectroscopic data
(This table is continued on the next page.) SPT ID of each cluster, right ascension and declination of its SZ center, and redshift z (from Tables 6.2 and 6.4, for reference). Also given are the SPT significance ξ and the SZ-based mass, marginalized over cosmological parameters as in Reichardt et al. (2013), for those clusters at $z \geq 0.3$. Clusters marked with ** are reported here as SPT detections for the first time, and those with * are new discoveries. References: (1) Sifón et al. (2012); (2) Girardi et al. (1996); (3) Struble & Rood (1999, this paper does not contain confidence intervals); (4) Barrena et al. (2002); (5) Katgert et al. (1998); (6) Buckley-Geer et al. (2011).

SPT ID	R.A. (J2000 deg.)	Dec. (J2000 deg.)	z	ξ	$M_{500c,SZ}$ ($10^{14} h_{70}^{-1} M_\odot$)	Source of spectro. this work	literature
SPT-CL J0000-5748	0.2387	−57.8063	0.702	7.71	4.32 ± 0.75	Y	
SPT-CL J0014-4952*	3.6913	−49.8729	0.752	8.90	5.46 ± 1.11	Y	
SPT-CL J0037-5047*	9.4382	−50.7938	1.026	6.94	4.07 ± 0.93	Y	
SPT-CL J0040-4407	10.2020	−44.1312	0.350	19.10	10.14 ± 2.00	Y	
SPT-CL J0102-4915	15.7271	−49.2562	0.870	39.58	15.54 ± 3.39		1
SPT-CL J0118-5156*	19.5932	−51.9396	0.705	5.59	3.50 ± 0.93	Y	
SPT-CL J0205-5829	31.4437	−58.4856	1.322	10.54	4.82 ± 0.96	Y	
SPT-CL J0205-6432	31.2786	−64.5461	0.744	6.02	3.39 ± 0.82	Y	
SPT-CL J0232-5257**	38.1813	−52.9562	0.556	8.83	5.46 ± 1.11		1
SPT-CL J0233-5819	38.2561	−58.3269	0.663	6.64	3.79 ± 0.86	Y	
SPT-CL J0234-5831	38.6790	−58.5217	0.415	14.65	7.71 ± 1.50	Y	
SPT-CL J0235-5120**	38.9443	−51.3479	0.278	9.28	-		1
SPT-CL J0236-4937**	39.2401	−49.6312	0.334	5.76	3.89 ± 0.96		1
SPT-CL J0240-5946	40.1620	−59.7703	0.400	9.04	5.32 ± 1.11	Y	
SPT-CL J0245-5302	41.3780	−53.0360	0.300	19.30	9.40 ± 3.02	Y	
SPT-CL J0254-5857	43.5729	−58.9526	0.437	14.42	7.61 ± 1.46	Y	
SPT-CL J0257-5732	44.3516	−57.5423	0.434	5.40	3.21 ± 0.86	Y	
SPT-CL J0304-4921**	46.0559	−49.3563	0.392	12.36	7.43 ± 1.43		1
SPT-CL J0317-5935	49.3208	−59.5856	0.469	5.91	3.54 ± 0.89	Y	
SPT-CL J0328-5541	52.1663	−55.6975	0.084	7.08	-		3
SPT-CL J0330-5227**	52.7237	−52.4646	0.442	11.38	6.36 ± 1.21		1
SPT-CL J0346-5438**	56.7210	−54.6479	0.530	8.42	4.86 ± 1.00		1
SPT-CL J0431-6126	67.8393	−61.4438	0.059	6.40	-		2
SPT-CL J0433-5630	68.2522	−56.5038	0.692	5.35	2.96 ± 0.82	Y	
SPT-CL J0438-5419	69.5686	−54.3187	0.422	22.24	10.43 ± 2.07	Y	1
SPT-CL J0449-4901*	72.2684	−49.0187	0.790	8.83	4.79 ± 0.93	Y	
SPT-CL J0509-5342	77.3360	−53.7045	0.462	6.61	5.36 ± 0.71	Y	1
SPT-CL J0511-5154	77.9202	−51.9044	0.645	5.63	3.71 ± 0.93	Y	
SPT-CL J0516-5430	79.1480	−54.5062	0.294	9.42	-	Y	
SPT-CL J0521-5104	80.2983	−51.0812	0.675	5.45	3.54 ± 0.96		1
SPT-CL J0528-5300	82.0173	−53.0001	0.769	5.45	3.21 ± 0.57	Y	1
SPT-CL J0533-5005	83.3984	−50.0918	0.881	5.59	2.75 ± 0.61	Y	
SPT-CL J0534-5937	83.6018	−59.6289	0.576	4.57	2.86 ± 1.00	Y	
SPT-CL J0546-5345	86.6541	−53.7615	1.066	7.69	5.29 ± 0.71	Y	1
SPT-CL J0551-5709	87.9016	−57.1565	0.424	6.13	3.82 ± 0.54	Y	
SPT-CL J0559-5249	89.9245	−52.8265	0.609	9.28	6.79 ± 0.86	Y	1
SPT-CL J0616-5227**	94.1393	−52.4562	0.684	9.29	5.64 ± 1.14		1
SPT-CL J0658-5556	104.6249	−55.9479	0.296	37.67	-		4

Table 3.1: SPT properties and source of spectroscopic data (continued)

SPT ID	R.A. (J2000 deg.)	Dec. (J2000 deg.)	z	ξ	$M_{500c,\mathrm{SZ}}$ ($10^{14} h_{70}^{-1} M_\odot$)	Source of spectro. this work	literature
SPT-CL J2012-5649	303.1132	−56.8308	0.055	5.99	-		2
SPT-CL J2022-6323	305.5235	−63.3973	0.383	6.58	3.89 ± 0.89	Y	
SPT-CL J2032-5627	308.0800	−56.4557	0.284	8.14	-	Y	
SPT-CL J2040-5725	310.0631	−57.4287	0.930	6.38	3.29 ± 0.79	Y	
SPT-CL J2043-5035	310.8285	−50.5929	0.723	7.81	4.79 ± 1.00	Y	
SPT-CL J2056-5459	314.2199	−54.9892	0.718	6.05	3.21 ± 0.79	Y	
SPT-CL J2058-5608	314.5893	−56.1454	0.606	5.02	2.71 ± 0.79	Y	
SPT-CL J2100-4548	315.0936	−45.8057	0.712	4.84	2.82 ± 0.89	Y	
SPT-CL J2104-5224	316.2283	−52.4044	0.799	5.32	3.14 ± 0.86	Y	
SPT-CL J2106-5844	316.5210	−58.7448	1.131	22.08	8.39 ± 1.68	Y	
SPT-CL J2118-5055	319.7291	−50.9329	0.625	5.62	3.54 ± 0.89	Y	
SPT-CL J2124-6124	321.1488	−61.4141	0.435	8.21	4.71 ± 1.00	Y	
SPT-CL J2130-6458	322.7285	−64.9764	0.316	7.57	4.54 ± 0.96	Y	
SPT-CL J2135-5726	323.9158	−57.4415	0.427	10.43	5.75 ± 1.11	Y	
SPT-CL J2136-4704	324.1175	−47.0803	0.425	6.17	4.11 ± 0.96	Y	
SPT-CL J2136-6307	324.2334	−63.1233	0.926	6.25	3.25 ± 0.75	Y	
SPT-CL J2138-6007	324.5060	−60.1324	0.319	12.64	6.82 ± 1.32	Y	
SPT-CL J2145-5644	326.4694	−56.7477	0.480	12.30	6.46 ± 1.25	Y	
SPT-CL J2146-4633	326.6473	−46.5505	0.931	9.59	5.43 ± 1.07	Y	
SPT-CL J2146-4846	326.5346	−48.7774	0.623	5.88	3.71 ± 0.93	Y	
SPT-CL J2148-6116	327.1798	−61.2791	0.571	7.27	4.11 ± 0.89	Y	
SPT-CL J2155-6048	328.9851	−60.8072	0.539	5.24	2.93 ± 0.79	Y	
SPT-CL J2201-5956	330.4727	−59.9473	0.098	13.99	-		5
SPT-CL J2248-4431	342.1859	−44.5271	0.351	40.97	17.29 ± 3.71	Y	
SPT-CL J2300-5331	345.1765	−53.5170	0.262	5.29	-	Y	
SPT-CL J2301-5546	345.4688	−55.7758	0.748	5.19	3.21 ± 0.93	Y	
SPT-CL J2325-4111	351.2985	−41.1937	0.358	12.27	7.36 ± 1.43	Y	
SPT-CL J2331-5051	352.9584	−50.8641	0.575	8.04	5.14 ± 0.71	Y	
SPT-CL J2332-5358	353.1040	−53.9733	0.402	7.30	6.54 ± 0.82	Y	
SPT-CL J2337-5942	354.3544	−59.7052	0.776	14.94	8.21 ± 1.14	Y	
SPT-CL J2341-5119	355.2994	−51.3328	1.002	9.65	5.61 ± 0.82	Y	
SPT-CL J2342-5411	355.6903	−54.1887	1.075	6.18	3.00 ± 0.50	Y	
SPT-CL J2344-4243	356.1817	−42.7229	0.595	27.53	12.39 ± 2.54	Y	
SPT-CL J2347-5158*	356.9423	−51.9766	0.869	4.36	2.32 ± 0.96	Y	
SPT-CL J2351-5452	357.8877	−54.8753	0.384	4.89	3.32 ± 1.04		6
SPT-CL J2355-5056	358.9551	−50.9367	0.320	5.89	4.11 ± 0.54	Y	
SPT-CL J2359-5009	359.9208	−50.1600	0.775	6.35	3.57 ± 0.57	Y	

3.3 Observations: optical spectroscopy

Instruments

The spectroscopic observations presented in this work are the first of our ongoing follow-up program. The data were taken from 2008 to 2012 using the Gemini Multi Object Spectrograph (GMOS; Hook et al., 2004) on Gemini South, the Focal Reducer

and low dispersion Spectrograph (FORS2; Appenzeller et al., 1998) on VLT Antu, the Inamori Magellan Areal Camera and Spectrograph (IMACS; Dressler et al., 2006) on Magellan Baade, and the Low Dispersion Survey Spectrograph (LDSS3[1]; Allington-Smith et al., 1994) on Magellan Clay.

In order to place a large number of slitlets in the central region of the cluster, most of the IMACS observations were conducted with the Gladders Image-Slicing Multi-slit Option (GISMO[2]). GISMO optically remaps the central region of the IMACS field-of-view (roughly $3.5' \times 3.2'$) to sixteen evenly-spaced regions of the focal plane, allowing for a large density of slitlets in the cluster core while minimizing slit collisions on the CCD. This is illustrated in Figure 3.1; the first stage of the remapping optics is visible in Figure 3.2, a photograph of the instrument.

[1]http://www.lco.cl/telescopes-information/magellan/instruments/ldss-3

[2]http://www.lco.cl/telescopes-information/magellan/instruments/imacs/gismo/gismoquickmanual.pdf

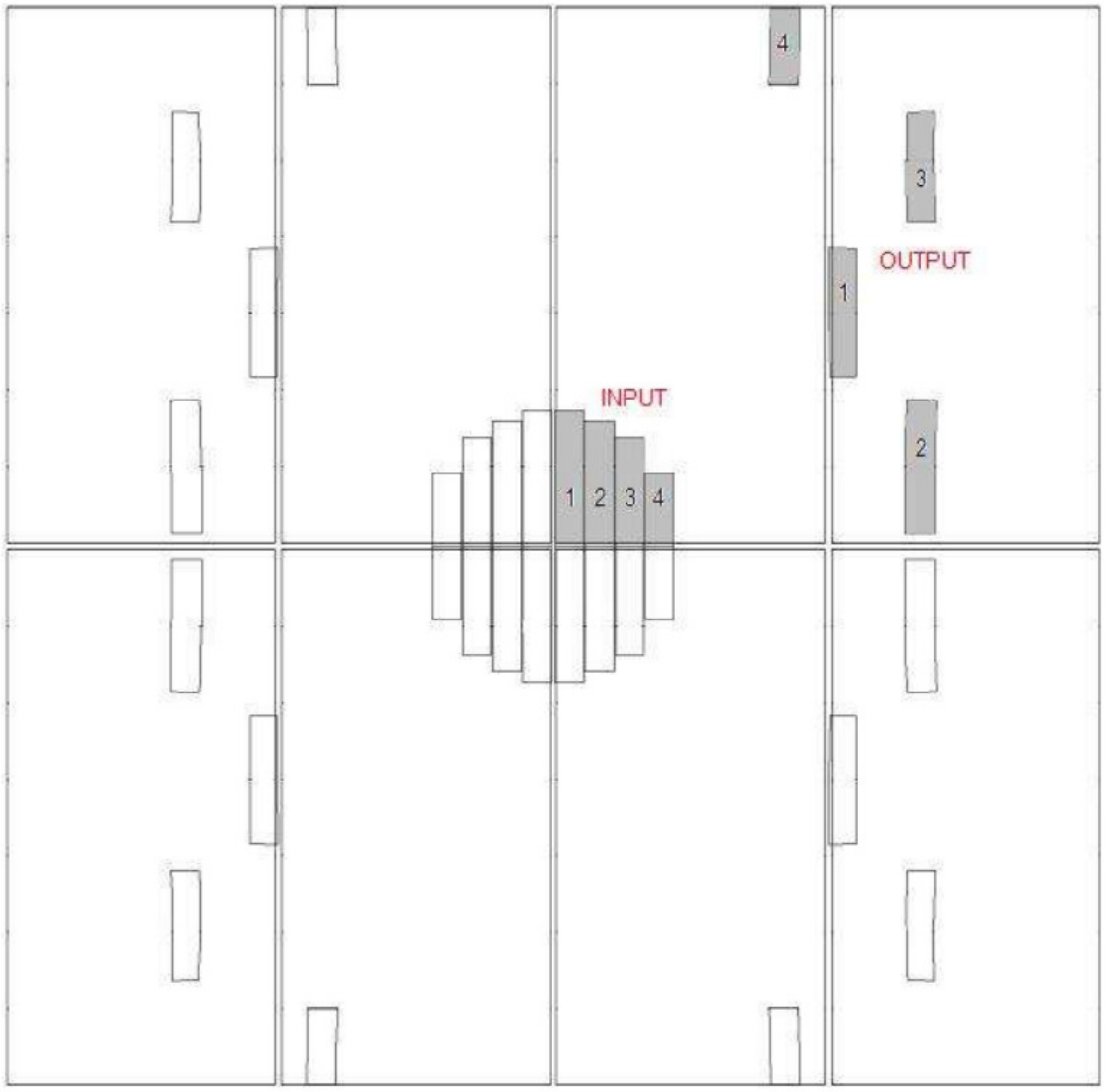

Figure 3.1: Footprint of GISMO input and output in the IMACS f/4 focal plane. The dispersion axis is left to right; compare this figure to the spectral traces in Figure 4.1. The field of view of the IMACS f/4 camera is a $15.46' \times 15.46'$ square, and the central non-rectangular region that is remapped has dimensions of roughly $3.5' \times 3.2'$. *Image credit: Mike Gladders/LCO, http://www.lco.cl/telescopes-information/magellan/instruments/imacs/gismo/gismoquickmanual.pdf*

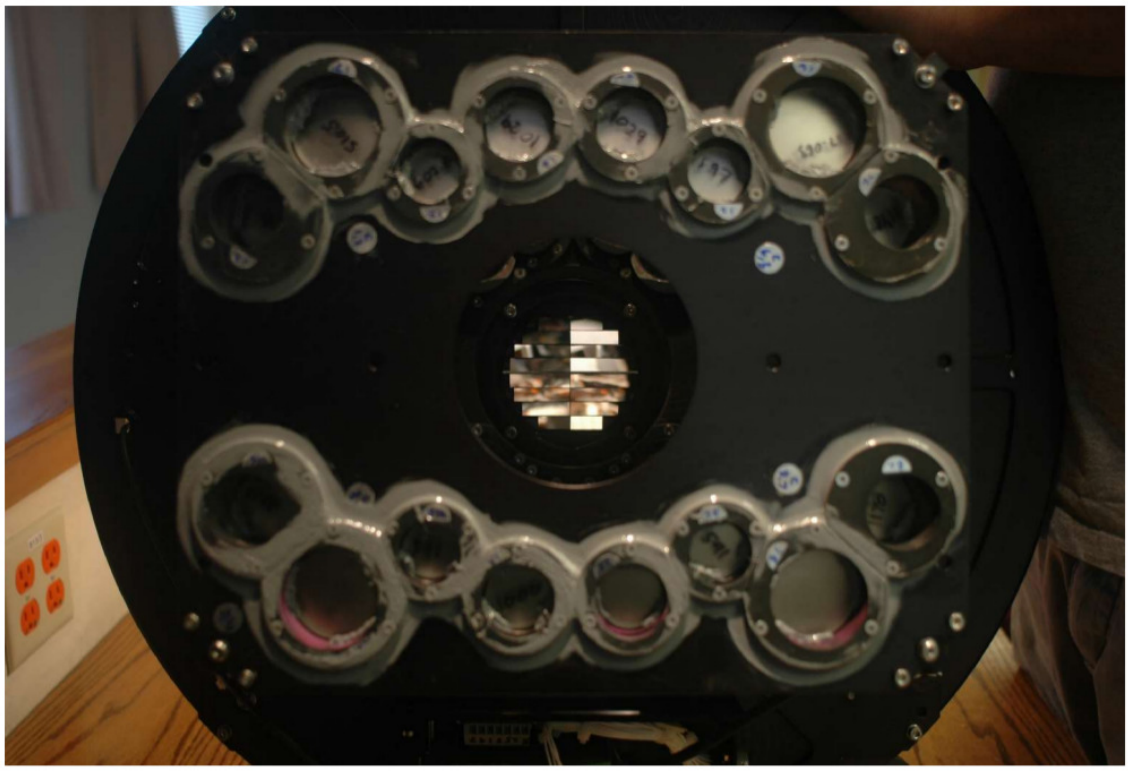

Figure 3.2: Photo of GISMO, where the slicing mirrors in the center are visible. *Image credit: Mike Gladders/LCO, http://www.lco.cl/telescopes-information/magellan/instruments/imacs/gismo/gismoquickmanual.pdf*

Upstream data

Optical and infrared follow-up imaging observations of SPT clusters are presented alongside our group's photometric redshift methodology in High et al. (2010) and Song et al. (2012). Those photometric redshifts (and in a few cases, spectroscopic redshifts from the literature) were used to guide the design of the spectroscopic observations. Multislit masks were designed using the best imaging available to us, usually a combination of ground-based *griz* (on Blanco/MOSAIC2, Magellan/IMACS, Magellan/LDSS3, or *BVRI* on Swope) and *Spitzer*/IRAC 3.6μm. In addition, spec-

troscopic observations at Gemini and VLT were preceded by single-band (r or i) pre-imaging for relative astrometry, or two-band (r and i) pre-imaging for red-sequence target selection in the cases where the existing imaging was not deep enough. The exposure times for this pre-imaging were chosen to reach a magnitude depth for galaxy photometry of $m^\star + 1$ at 10σ at the cluster redshift.

Mask design

In designing the multislit masks, top priority for slit placement was given to bright red-sequence galaxies (the red sequence of SPT clusters is discussed in the context of photometric redshifts in High et al., 2010; Song et al., 2012), as defined by their distance to either a theoretical or an empirically-fit red-sequence model. The details varied depending on the quality of the available imaging, the program and the prioritization weighting scheme of the instrument's mask-making software. In many of the GISMO observations and some of the Gemini observations, blue galaxies were given higher priority than faint red galaxies because, especially at high redshift, they were expected to be more likely to yield a redshift. An example of color cuts used in mask design is shown in Figure 3.3.

The results from the different red-sequence weighting schemes are very similar, and few emission lines are found, even at high redshift ($z > 1$; Brodwin et al., 2010; Foley et al., 2011; Stalder et al., 2013, these articles also provide more details about the red-sequence nature of spectroscopic members). In all cases, non-red-sequence objects were used to fill out any remaining space in the mask. Figure 3.4 shows the same color-magnitude diagram as Figure 3.3, where the spectroscopic members

from the observation of one multislit mask are shown in red, and the spectroscopic non-members are shown in green.

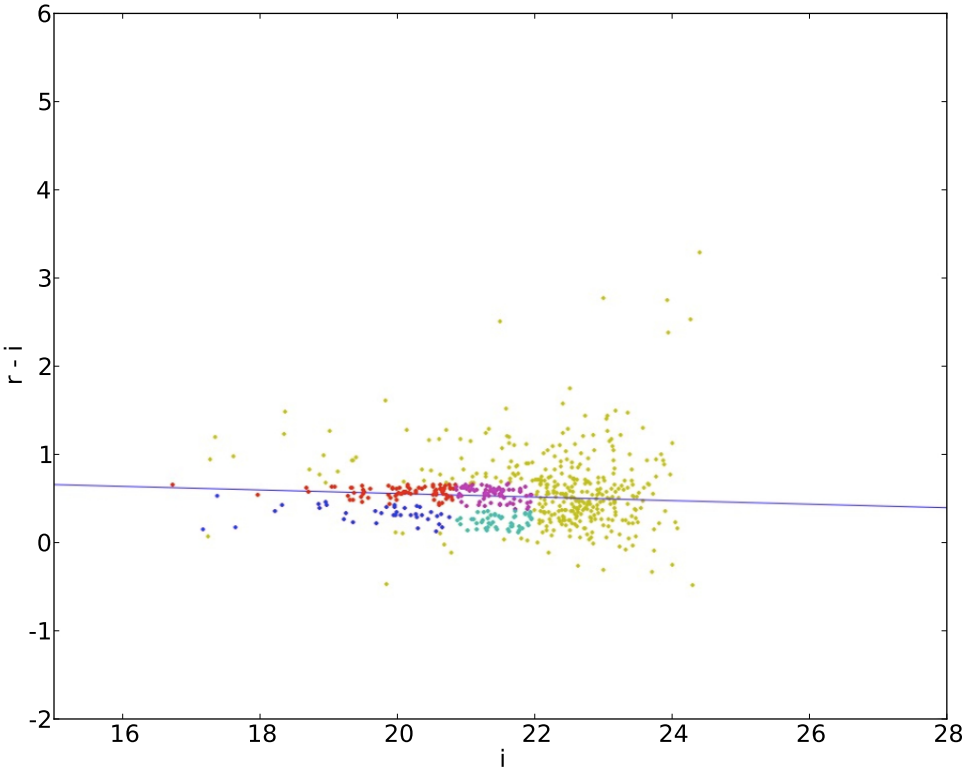

Figure 3.3: Color-Magnitude Diagram of objects in the SPT-CLJ0438-5419 field-of-view. The straight line shows an empirical red-sequence, and the colors classes of objects given different weights in the design of spectroscopic masks. The classes are bright red (red), bright blue (blue), faint blue (teal) and faint red (purple). An optical image of SPT-CLJ0438-5419 is found in Figure 1.1

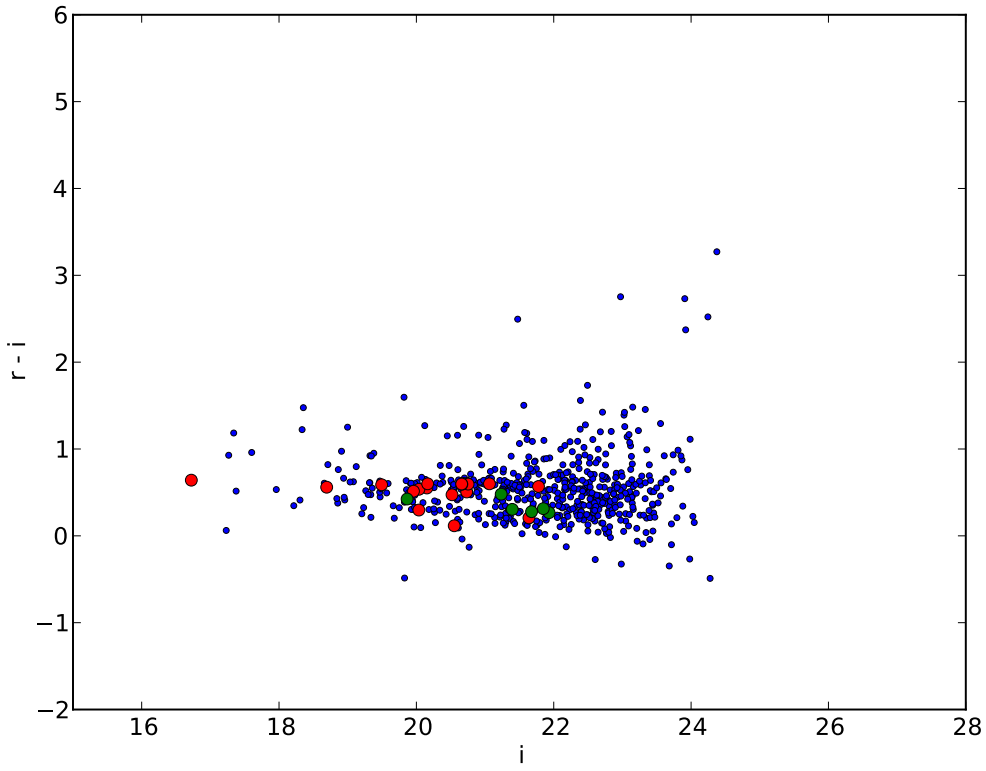

Figure 3.4: Color-Magnitude Diagram of objects in the SPT-CLJ0438-5419 field-of-view. The spectroscopic members and non-members from the observation of one mask are shown in red and green, respectively. The central galaxy (and BCG, brightest cluster galaxy) is the leftmost red point. An optical image of SPT-CLJ0438-5419 is found in Figure 1.1

Detail of observations

Details about the observations pertaining to each cluster, including the instrument, optical configuration, number of masks, total exposure time, and measured spectral resolution are listed in Table 3.2.

The dispersers and filters, listed in Table 3.2, were chosen (within the uncertainty in the photo-z) to obtain low- to medium-resolution spectra covering at least the

wavelengths of the main spectral features that we use to identify the galaxy redshifts: [O II] emission, and the Ca II H&K absorption lines and break.

The spectroscopic exposure times (also in Table 3.2) for GMOS and FORS2 observations were chosen to reach $S/N = 5$ ($S/N = 3$) per spectral element just below the 4000Å break for a red galaxy of magnitude $m^\star + 1$ ($m^\star + 0.5$) at $z < 1$ ($z > 1$). Under the observing conditions prevailing at the telescope during classical observing, the exposure times for the Magellan observations were determined by a combination of experience, real-time quick-look reductions, and airmass limitations.

Table 3.2: Observations

(*This table is continued on next page.*) The instruments used for our observations are IMACS on Magellan Baade, LDSS3 on Magellan Clay, GMOS-S on Gemini South, and FORS2 on VLT Antu. The UT date of observation, details of the configuration and the number of observed multislit masks are given, as well as the number of member redshifts retrieved from the observation ($N \equiv N_{\rm members}$), and the total spectroscopic exposure time for all masks, $t_{\rm exp}$, in hours. The spectral resolution R is the FWHM of sky lines in Angstroms, measured in the science exposures.

† In these observations, the cluster members were secondary targets.

SPT ID	z	UT Date	Instrument	Disperser/Filter	Masks	N	$t_{\rm exp}$ (h)	R (Å)
SPT-CL J0000-5748	0.702	2010 Sep 07	GMOS-S	R150_G5326	2	26	1.33	23.7
SPT-CL J0014-4952	0.752	2011 Aug 21	FORS2	GRIS_300I/OG590	2	29	2.83	13.5
SPT-CL J0037-5047	1.026	2011 Aug 22	FORS2	GRIS_300I/OG590	2	18	5.00	13.5
SPT-CL J0040-4407	0.350	2011 Sep 29	GMOS-S	B600_G5323	2	36	1.17	5.7
SPT-CL J0118-5156	0.705	2011 Sep 28	GMOS-S	R400_G5325, N&S	2	14	2.53	9.0
SPT-CL J0205-5829	1.322	2011 Sep 25	IMACS	Gri-300-26.7/WB6300-950, f/2	1	9	11.00	5.2
SPT-CL J0205-6432	0.744	2011 Sep 30	GMOS-S	R400_G5325, N&S	2	15	2.67	9.0
SPT-CL J0233-5819	0.664	2011 Sep 29	GMOS-S	R400_G5325, N&S	1	10	1.33	9.0
SPT-CL J0234-5831	0.415	2010 Oct 08	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	22	1.50	6.5
SPT-CL J0240-5946	0.400	2010 Oct 09	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	25	1.00	6.4
SPT-CL J0245-5302	0.300	2011 Sep 29	GMOS-S	B600_G5323	2	29	0.83	7.0
SPT-CL J0254-5857	0.437	2010 Oct 08	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	35	1.50	6.9
SPT-CL J0257-5732	0.434	2010 Oct 09	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	22	1.50	6.6
SPT-CL J0317-5935	0.469	2010 Oct 09	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	17	1.63	6.6
SPT-CL J0433-5630	0.692	2011 Jan 28	IMACS/GISMO	Gri-300-17.5/Z2-520-775, f/2	1	22	1.00	5.7
SPT-CL J0438-5419	0.422	2011 Sep 28	GMOS-S	R400_G5325	1	18	0.75	9.0
SPT-CL J0449-4901	0.790	2011 Jan 28	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	20	1.63	5.6
SPT-CL J0509-5342	0.462	2009 Dec 12	GMOS-S	R150_G5326	2	18	1.00	23.7
		2012 Mar 23	FORS2	GRIS_300V/GG435	1	4†	2.37	13.7
SPT-CL J0511-5154	0.645	2011 Sep 30	GMOS-S	R400_G5325, N&S	2	15	2.67	9.0
SPT-CL J0516-5430	0.294	2010 Sep 17	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	2	48	1.67	6.7
SPT-CL J0528-5300	0.769	2010 Jan 13	GMOS-S	R150_G5326	2	20	3.00	23.7
SPT-CL J0533-5005	0.881	2008 Dec 05	LDSS3	VPH-Red	1	4	0.63	5.4
SPT-CL J0534-5937	0.576	2008 Dec 05	LDSS3	VPH-Red	1	3	0.45	5.5
SPT-CL J0546-5345	1.066	2010 Feb 11	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	21	3.00	5.7
SPT-CL J0551-5709	0.424	2010 Sep 17	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	2	34	1.42	6.8
SPT-CL J0559-5249	0.609	2009 Dec 07	GMOS-S	R150_G5326	2	37	1.33	23.7
SPT-CL J2022-6323	0.383	2010 Oct 09	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	37	1.17	6.7
SPT-CL J2032-5627	0.284	2010 Oct 08	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	31	1.17	6.8
SPT-CL J2040-5725	0.930	2010 Aug 13	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	5	3.00	5.0

Table 3.2: Observations (continued)

SPT ID	z	UT Date	Instrument	Disperser/Filter	Masks	N	$t_{\rm exp}$ (h)	R (Å)
SPT-CL J2043-5035	0.723	2011 Aug 27	FORS2	GRIS_300I/OG590	2	21	4.00	13.5
SPT-CL J2056-5459	0.719	2010 Aug 14	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	12	2.00	5.3
SPT-CL J2058-5608	0.607	2011 Oct 01	GMOS-S	R400_G5325	2	9	1.67	9.0
SPT-CL J2100-4548	0.712	2011 Jul 23	FORS2	GRIS_300I/OG590	2	19	1.50	13.5
SPT-CL J2104-5224	0.799	2011 Jul 21	FORS2	GRIS_300I/OG590	2	23	2.83	13.5
SPT-CL J2106-5844	1.131	2010 Dec 08	FORS2	GRIS_300I/OG590	1	15	3.00	13.5
		2010 Jun 07	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	4	8.00	4.5
SPT-CL J2118-5055	0.625	2011 May 26	FORS2	GRIS_300I/OG590	2	22	1.33	13.5
		2011 Sep 27	GMOS-S	R400_G5325, N&S	1	3	1.20	9.0
SPT-CL J2124-6124	0.435	2009 Sep 25	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	24	1.50	7.0
SPT-CL J2130-6458	0.316	2010 Sep 17	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	2	47	2.00	7.1
SPT-CL J2135-5726	0.427	2010 Sep 16	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	33	1.00	6.8
SPT-CL J2136-4704	0.425	2011 Sep 29	GMOS-S	R400_G5325	2	24	1.67	9.0
SPT-CL J2136-6307	0.926	2010 Aug 14	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	10	2.00	5.0
SPT-CL J2138-6007	0.319	2010 Sep 17	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	34	1.50	6.8
SPT-CL J2145-5644	0.480	2010 Sep 16	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	2	37	2.92	7.4
SPT-CL J2146-4633	0.931	2011 Sep 25	IMACS	Gri-300-26.7/WB6300-950, f/2	1	17	3.00	4.7
SPT-CL J2146-4846	0.623	2011 Oct 01	GMOS-S	R400_G5325	2	26	2.33	9.0
SPT-CL J2148-6116	0.571	2009 Sep 25	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	30	1.50	7.1
SPT-CL J2155-6048	0.539	2011 Oct 01	GMOS-S	R400_G5325	2	25	1.50	9.0
SPT-CL J2248-4431	0.351	2009 Jul 12	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	15	1.33	10.9
SPT-CL J2300-5331	0.262	2010 Oct 08	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	24	1.00	6.8
SPT-CL J2301-5546	0.748	2010 Aug 14	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	11	2.00	5.4
SPT-CL J2325-4111	0.358	2011 Sep 28	GMOS-S	B600_G5323	2	33	1.00	5.7
SPT-CL J2331-5051	0.575	2008 Dec 05	LDSS3	VPH-Red	2	6	1.00	5.5
		2010 Sep 09	GMOS-S	R150_G5326	2	28	1.00	23.7
		2010 Oct 09	IMACS/GISMO	Gra-300-4.3/Z2-520-775, f/4	2	62	3.50	6.7
SPT-CL J2332-5358	0.403	2009 Jul 12	IMACS/GISMO	Gri-200-15.0/WB5694-9819, f/2	1	24	1.50	18.1
		2010 Sep 05	FORS2	GRIS_300V	2	29†	4.38	13.7
SPT-CL J2337-5942	0.776	2010 Aug 14	GMOS-S	R150_G5326	2	19	3.00	23.7
SPT-CL J2341-5119	1.003	2010 Aug 14	GMOS-S	R150_G5326	2	15	6.00	23.7
SPT-CL J2342-5411	1.075	2010 Sep 09	GMOS-S	R150_G5326	1	11	3.00	23.7
SPT-CL J2344-4243	0.595	2011 Sep 30	GMOS-S	R400_G5325	2	32	2.33	9.0
SPT-CL J2347-5158	0.869	2010 Aug 13	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	12	2.50	5.0
SPT-CL J2355-5056	0.320	2010 Sep 17	IMACS/GISMO	Gra-300-4.3/Z1-430-675, f/4	1	37	1.50	7.0
SPT-CL J2359-5009	0.775	2009 Nov 22	GMOS-S	R150_G5326	2	7	1.33	23.7
		2010 Aug 14	IMACS/GISMO	Gri-300-26.7/WB6300-950, f/2	1	22	2.00	5.4

Chapter 4

Spectroscopic data processing

In this chapter, we present a quick review of the different steps of the data processing, with a special mention of details specific to multi-object spectroscopy, and the use of the GISMO instrument. It is also our hope that the figures in particular will provide the reader with a more concrete sense of the data that produced the numbers presented in Chapter 6.

Most of the following explanations apply to general spectroscopic CCD observations; however, some details would be different for instruments other than the ones that we have used (IMACS/GISMO, GMOS and FORS2; see 3.3), most notably fiber-fed spectrographs, which we will not discuss as we have not used them.

A general introduction and reference for the astronomical use of CCDs is Howell (2006).

4.1 CCD reductions

In a spectroscopic observation, the light from the galaxies which we want to observe is dispersed by a grating or grism (a combination of a grating and prism), and the dispersed trace is imaged by a CCD. To create a dispersed image that is easy to interpret and process, a mask milled from an aluminum plate is inserted into the telescope beam and precisely aligned on the sky to block the light from the sky and from all objects in the camera's field of view, except for a number of slits that let the light of the target galaxies through, as well as some of the sky to allow for determination – and subtraction – of the sky spectrum. In the image, each slit produces a rectangle where the short axis is a spatial coordinate on the sky, and the long axis is mapped to the wavelength. The object spectrum is spatially localized at the center of the slit, and superimposed on the sky spectrum which has no spatial dependence. See the bottom image of Figure 4.2 for an exemple.

The CCD is an array of silicon pixels; during an exposure, photons create free electron-hole pairs in the silicon that remain trapped in the pixel until readout. During readout, the pixel values are read sequentially, pixel by pixel, column by column. The readout amplifier converts the voltage from the electrons in each pixel to a digital number, in ADUs, or analog-to-digital units. The number of electrons corresponding to an ADU is called the *gain*. It has units of electrons per ADU.

For the CCD reductions described in this section, we used the COSMOS reduction package[1] (Kelson, 2003) for CCD reductions of IMACS and LDSS3 data, and standard

[1] http://code.obs.carnegiescience.edu/cosmos

IRAF routines and XIDL[2] routines for GMOS and FORS2. Of those packages, the author has the most experience with COSMOS.

Bias subtraction and flat-fielding

At readout time, a bias level is added to the counts in ADUs to avoid negative count values. This bias varies spatially across the CCD and can be measured by taking a number of *bias frames*, or zero-second exposures. The afternoon calibrations at the telescope usually involve taking at least 10 such frames. Their average (with rejection of high values, as there can be cosmic rays even in bias frames) is the bias level to be subtracted from all other types of exposures. An example of a bias frame is shown as the first element of Figures 4.1 and 4.2.

The other essential calibration frame type for both imaging and spectroscopy is a *flat frame*, for flat-fielding. Flat-fielding serves to correct the response of each pixel to a reference illumination. This response is dependent on wavelength and so different flat frames must be used for different filters, and in the case of spectroscopy, different dispersers and masks, which will change the wavelength solution (see below).

The flat frames are taken by exposing a uniform light to a high number of counts per pixel. In the case of spectroscopy, a smooth 2-dimensional illumination function is fit to each slit. If we call the number of counts of this smooth fit the model, then the response of each pixel is actual counts divided by model. Flat-fielding means dividing the counts by the response.

The strictest, most basic meaning of CCD data reduction is bias subtraction and

[2]http://www.ucolick.org/~xavier/IDL/

flat-fielding, which produce counts in units of ADUs. When the detector is a mosaic of multiple chips, it is also necessary to correct for the gain, which varies chip by chip, to convert the counts to units of electrons. These steps have to be done for both imaging and spectroscopy.

Schematically, the basic reduction is (exposure - bias)/response.

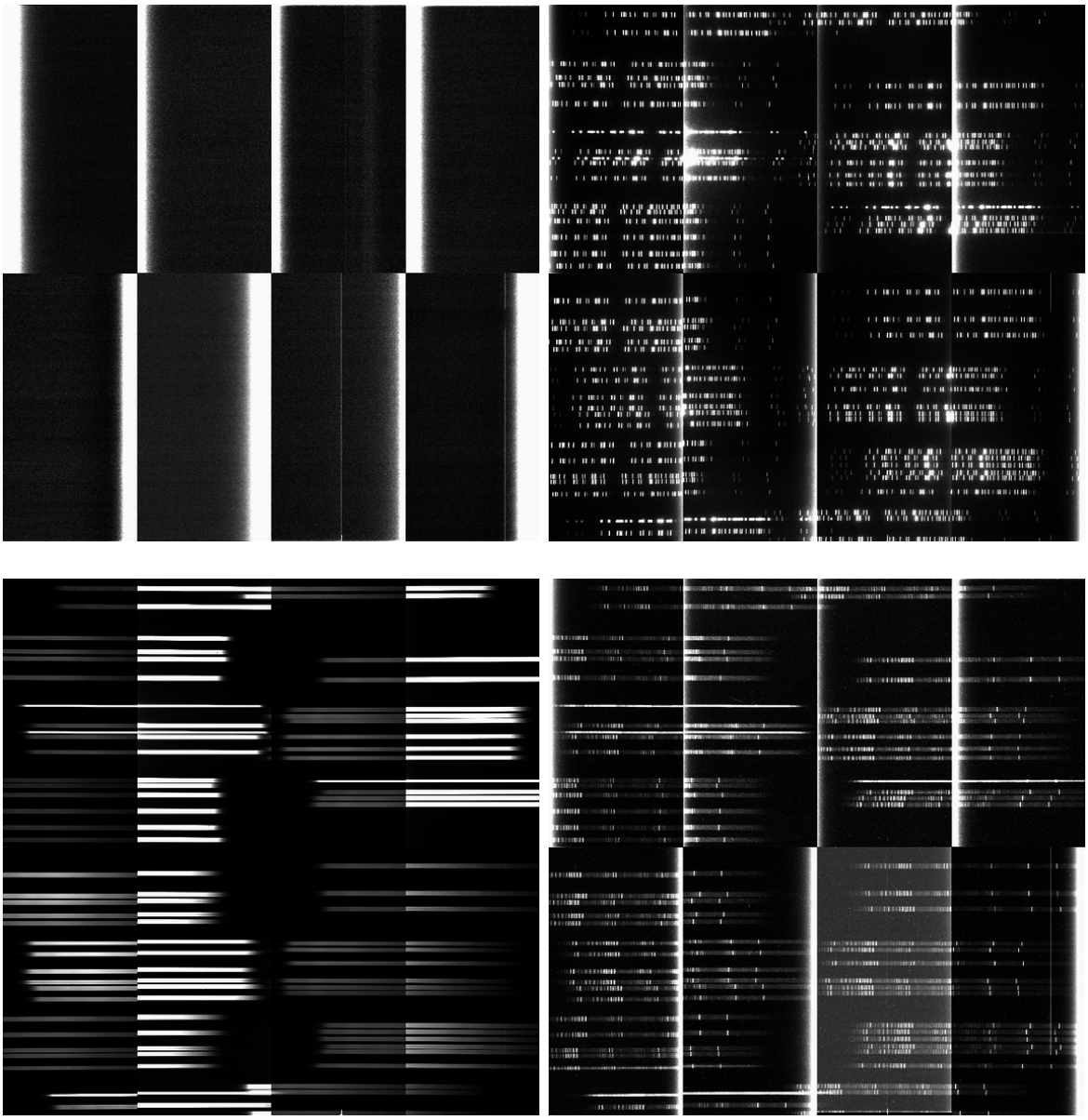

Figure 4.1: Raw exposures from the IMACS f/4 camera. The camera has eight CCD chips, hence each exposure is a mosaic of eight chips. There are small chipgaps not shown here. The top left mosaic is a bias frame, the top right an arc frame, the bottom left is a spectroscopic flat, and the bottom right is a science exposure. A subsection of these exposures is shown with greater magnification in Figure 4.2

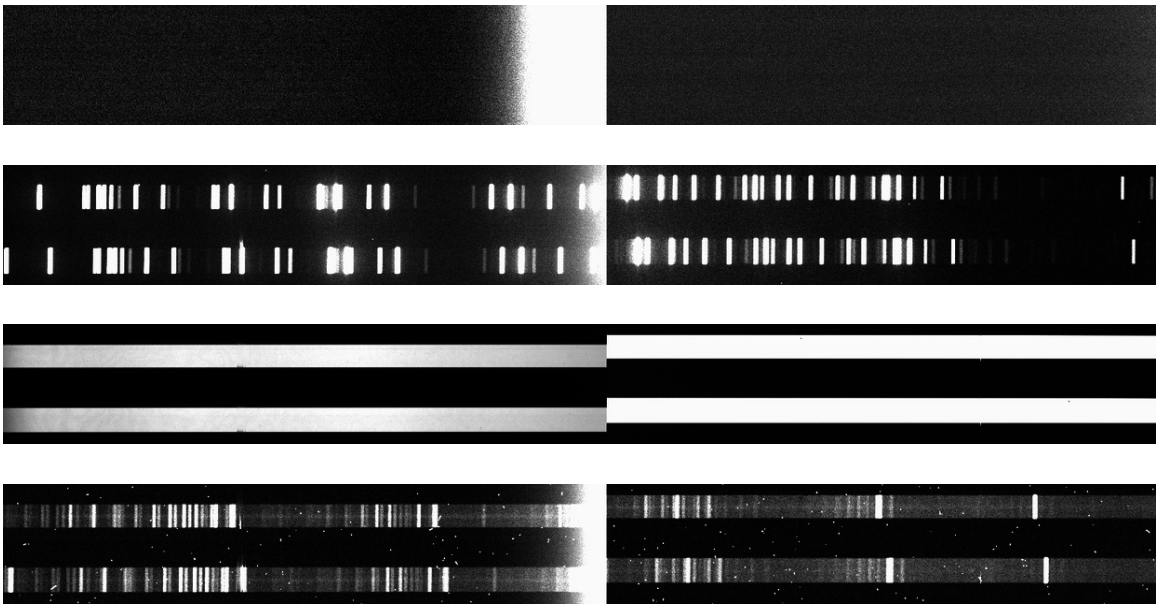

Figure 4.2: Detail of raw exposures from the IMACS f/4 camera; these are the same exposures as shown in Figure 4.1. From top to bottom, these are a bias frame, an arc frame, a spectroscopic flat, and a science exposure. The discontinuity in the center is a chip gap.

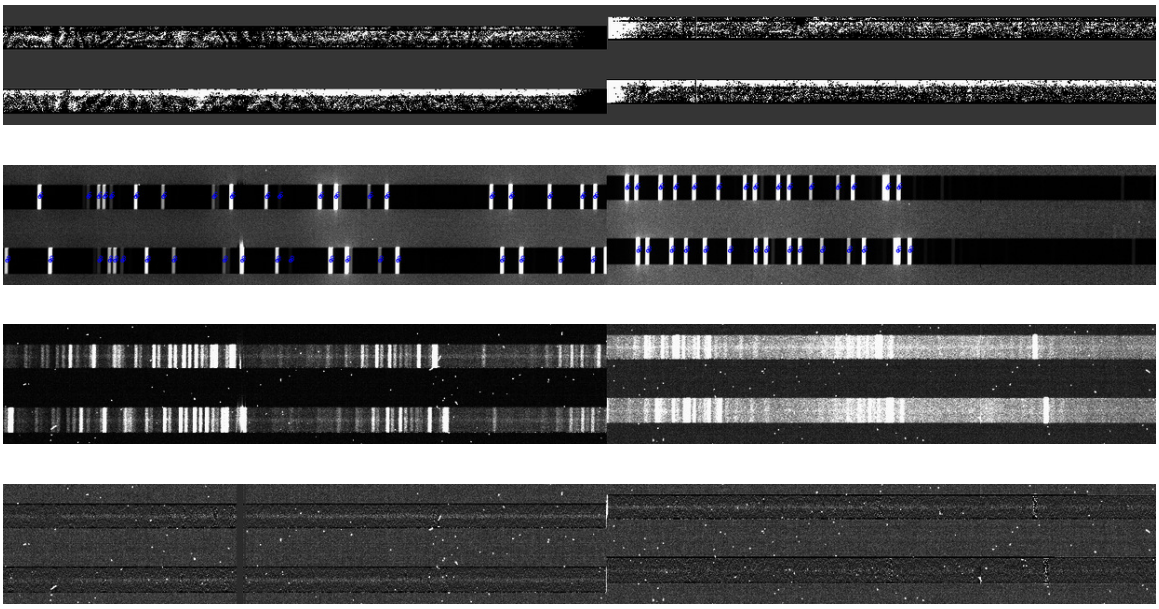

Figure 4.3: Detail of processed exposures from the IMACS f/4 camera; these are the same traces as shown in Figure 4.2. From top to bottom, these are 1) the response function for flat-fielding, as computed from the flat frame, 2) a bias-subtracted, flat-fielded arc-frame showing the location of the lines as computed using the wavelenght solution, 3) a bias-subtracted, flat-fielded science exposure, and 4) the same science exposure after sky subtraction.

Wavelength calibration

Wavelength calibration is based on arc lamp exposures, obtained at night in between science exposures in the case of IMACS and LDSS3, and during day time in the same configuration as for science exposures for GMOS and FORS2. In the case of day-time arc frames, the wavelength calibration was refined using sky lines in the science exposures.

Typically, the trace of each slit is aligned so that there is a rough correspondance between e.g. the rows or the x axis of the pixels and the wavelength, and the columns or the y axis and the spatial direction along the slit, but small distortions really

make it such that the wavelength is a function of both x and y. Getting a good two-dimensional wavelength solution is important not only for doing the science that depends on the wavelength, but also the sky subtraction (see below).

The distortion in the slit is mapped in x and y by cross-correlating rows with one another in an arc frame, globally and in wavelength segments, from which a remapped slit where the new x' and y' truly correspond to wavelength and spatial position can be constructed. The wavelength can then be calibrated to nanometers because the arc lines are easy to identify and have known wavelengths; for instance, they are Helium, Neon and Argon lines for IMACS.

The combination of distortion and wavelength solution is typically well fit by a third-order polynomial in x and y for IMACS data.

The second image of Figure 4.3 shows the detail of an arc frame where the fit location of lines is superimposed on the exposure (as does the first image of Figure 4.4, in a slightly more complicated situation of trace overlap).

Sky subtraction

We use the "optimal" sky substraction algorithm of Kelson (2003).

The simplest sky-subtraction algorithm that we could imagine would be to use the distortion map and wavelength solution to resample the image and produce an image where the coordinates are y vs λ. We call this step rectification. One can then define some range or ranges of y to be the sky, and then produce a model of the sky as a function of lambda, and subtract it everywhere.

The Kelson (2003) algorithm does the sky subtraction before rectification, which

leaves fewer artifacts in the sky-subtracted, rectified 2d spectrum than the simpler procedure just outlined. High pixels are rejected in a robust way that looks at a few-column average at a time, so as to exclude both the object spectrum and cosmic rays, but not bright sky lines. After the rejection, one is left with the "sky pixels".

The distortion map yields a wavelength for each pixel, and the sky spectrum is spline-interpolated from the λ, flux pairs of the sky pixels.

This sky spectrum is then resampled at the λs of all the pixels, to create the two-dimensional sky model that is subtracted.

The bottom two images of Figure 4.3 shows a detail of a bias-subtracted, flat-fielded science exposure before sky subtraction (third image from the top) and after sky subtraction (bottom image). The bottom image of Figure 4.4 also shows a sky-subtracted science frame in a slightly more complicated situation of trace overlap.

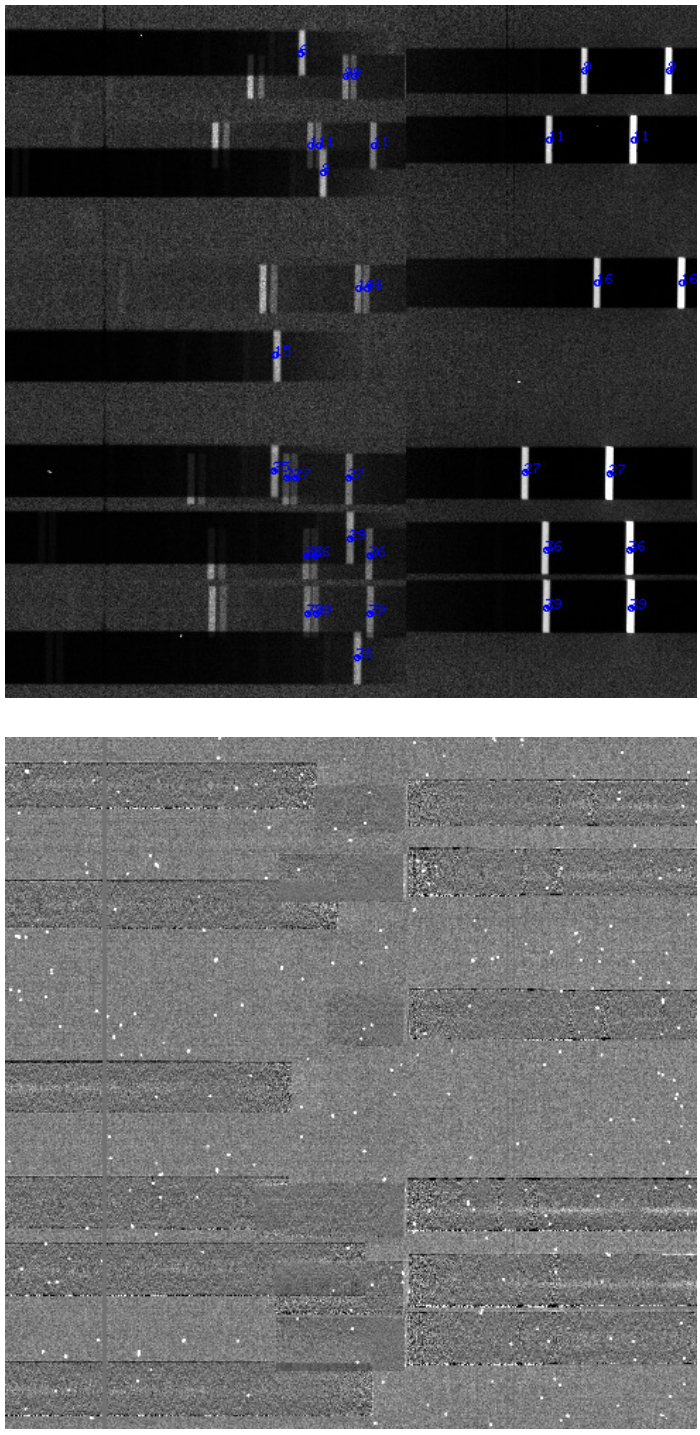

Figure 4.4: Overlap of slits on the CCD. The top panel shows the wavelength solution, and the bottom panel a sky-subtracted science exposure.

Considerations specific to multi-object spectroscopy

For the most part, mutli-object spectroscopy is not different from single-slit or long slit spectroscopy, as each slit can be reduced independently.

An added complication is that of trace collisions, where either 1) the trace from two different slits overlap or 2) in the case of grating dispersion, zeroth-order light from a slit overlaps with another slit (the first-order, dispersed slit proper), potentially causing problems in the wavelength solution and sky subtraction.

Trace collisions should not happen often in well-designed masks, however the extent of the trace in the wavelength direction depends on the filter response and intensity of the light, so can become a potential issue in arc frames and flat frames, which are very bright.

For our Gemini observations, any row on the CCD had at most one slit (in other words the slits were stacked vertically) so this was not a problem. It was a problem, however, in the case of GISMO because of the remapping.

Figure 4.4 shows a region of an IMACS / GISMO exposure where slit ends overlap.

In principle, the reduction software could be made to recognize and exclude such regions of overlap. The COSMOS software does not do that, and the volume of data in this study was manageable enough for the author to verify each wavelength solution fit by eye. The wavelength fits were largely unaffected, and any failures could be fixed by rejection of problematic lines from the fit. The first image of Figure 4.4 shows the location of lines computed from the wavelength solution, overplotted on an arc frame.

Even with a good wavelength solution, the sky subtraction was still problematic in collisions, given the algorithm used by COSMOS. This too could have been noted

in the software in a well-integrated pipeline. In the present case, the subsequent extraction and analysis of the extracted spectrum were always done with inspection of the 2d spectrum, so in the small number of occurences of bad sky subtraction due to collisions, the bad regions were excluded from fits manually. The second image of Figure 4.4 shows the sky-subtracted collision region.

4.2 Extracted spectrum processing and cross-correlation for velocity

Flux calibration

Flux calibration and telluric line removal were performed using the well-exposed continua of spectrophotometric standard stars (Wade & Horne, 1988; Foley et al., 2003). On the night of the observations, a standard star observable at a similar airmass to the science targets was observed using the same optical configuration as the science observations. The processed, extracted 1d spectrum flux level is compared to tabulated values, so that the relative response of the instrument as a function of wavelength can be inferred and modeled via a spline fit. The amount of sky absorption in regions of the spectrum where it is known to happen is then computed by the difference between the spline-fit continuum level and the measured level.

Cross-correlation for velocity

Fitting spectral features with known rest-frame wavelengths to the observed spectrum yields a redshift z_i for each galaxy, via

$$\lambda_{\text{observed}} = \lambda_{\text{rest}}(1 + z_i). \tag{4.1}$$

The redshift determination was performed by the author using cross-correlation with the *fabtemp97* template in the RVSAO package (Kurtz & Mink, 1998) for IRAF or by a collaborator with an in-house template fitting method using the SDSS DR2 templates. In all cases the fits were validated by agreement with visually identified absorption or emission features. A single method was used for each cluster depending on the workflow, and both perform similarly. Comparison between those redshifts obtained from the continuum and emission-line redshifts, when both are available from the same spectrum, shows that the uncertainties in individual redshifts (twice the RVSAO uncertainty, see e.g. Quintana et al., 2000) correctly represent the statistical uncertainty of the fit.

RVSAO is a collection of several different routines, of which we used *xcsao* for determining the redshift from absorption features, and *emsao* for emission lines. Screen grabs of both routines are shown in Figures 4.5 and 4.6.

Because we had the criterion that the redshift determination needed to be confirmed by eye by an identifiable absorption or emission feature, the accuracy of the automatic fitting offered by those routines is not an issue that needs to be considered.

xcsao fits and subtracts the continuum from the spectrum and uses cross-correlation with an absorption-line template to determine the redshift. There is information lost in subtracting the continuum, but it also eliminates the requirement of a good flux

calibration. Given our requirement of a visual confirmation of the redshift, cases where we would have had a low-confidence fit from the continuum only were discarded. Also, comparison with the template fitting method showed no significant difference in the ability to retrieve redshifts.

xcsao was run interactively, so that the correct local maximum of the cross-correlation could be picked manually. Also, one of the user-defined parameters is the wavelength range within which the cross-correlation is calculated. This range was adjusted to exclude any regions of bad sky subtraction, be it due to trace overlaps, or other reasons like zeroth-order light.

emsao, like *xcsao*, fits and subtracts the continuum. Gaussian profiles are fit to emission lines, and running the routine interactively allows to exclude or add lines to those found automatically.

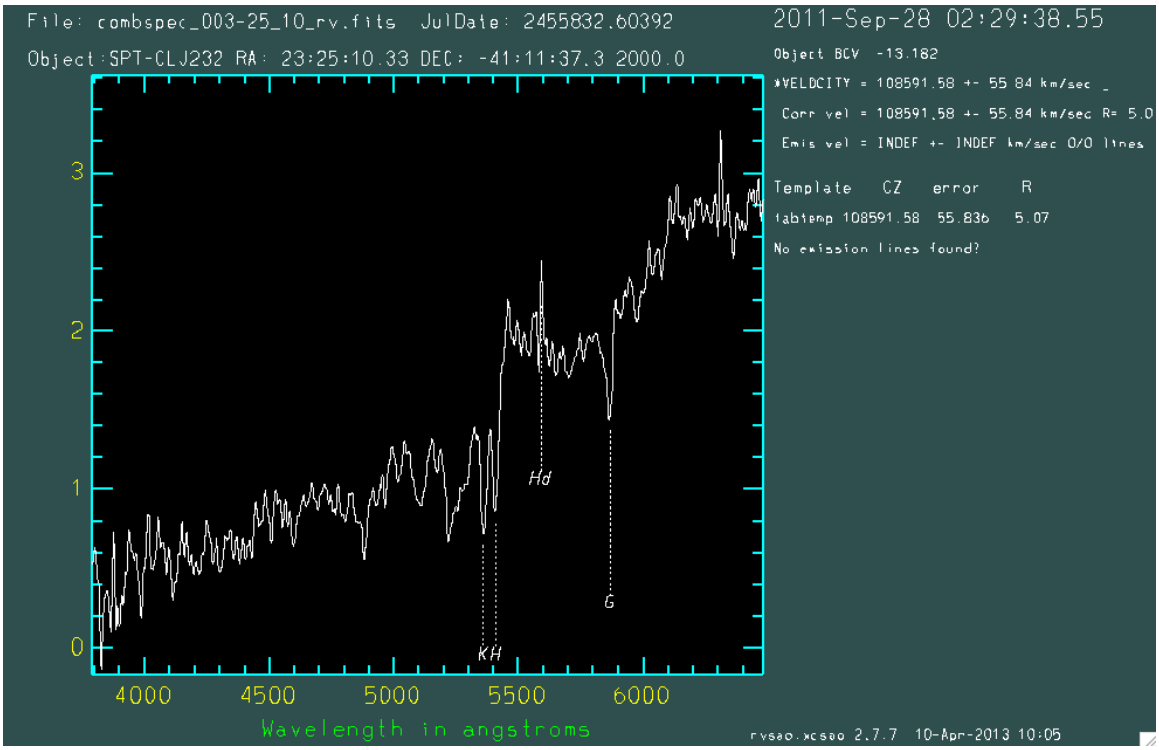

Figure 4.5: Screen grab of the *xcsao* routine showing the cross-correlation fit to a red galaxy spectrum by labeling the absorption lines. The Ca II H&K lines just below a break in the continuum level makes this an unambiguous identification. Most red-sequence galaxies with a redshift in our sample had their redshift identified from these same features.

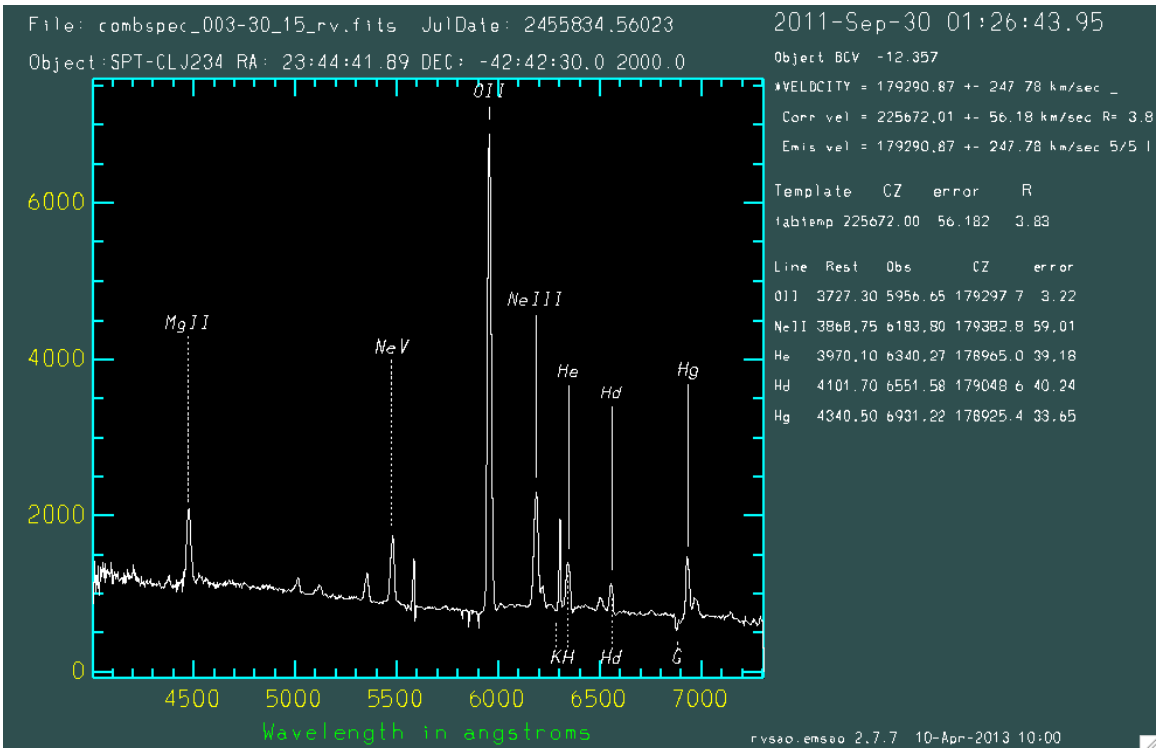

Figure 4.6: Screen grab illustrating the use of the *emsao* routine to fit a redshift to the extraordinary emission spectrum of the central galaxy of SPT-CL J2344-4243, a very massive SPT cluster at $z = 0.595$. This particular cluster is studied in McDonald et al. (2012).

Chapter 5

Methods and statistics of velocity dispersion measurements

In this chapter, we review and develop the different mathematical steps that allow us to go from the measured redshifts of individual galaxies to a cluster redshift and cluster velocity dispersion with appropriate uncertainties. We review methods that are *de facto* standard in the velocity dispersion literature in Section 5.1, and then move on to using our data (to be presented in Chapter 6) for a statistical exploration using resampling, in Sections 5.2 and 5.3.

5.1 Standard methods

5.1.1 From redshifts to peculiar velocities

Fitting spectral features with known rest-frame wavelengths to the observed spectrum yields a redshift z_i for each galaxy, via

$$\lambda_{\rm observed} = \lambda_{\rm rest}(1 + z_i). \tag{5.1}$$

This observed redshift has contributions from three different sources: the Doppler shift due to the peculiar velocity of the Earth with respect to the CMB $(1 + z_\oplus)$, the contribution from the cosmological redshift of the cluster $(1 + \bar{z})$, and the Doppler shift due to the peculiar velocity of the galaxy within the cluster, $(1 + z_{{\rm p},i})$:

$$(1 + z_i) = (1 + z_\oplus)(1 + \bar{z})(1 + z_{{\rm p},i}). \tag{5.2}$$

In theory, the cluster also has a proper velocity with respect to its cosmological comoving frame, but we omit it because in practice we cannot separate its contribution from the cosmological redshift. Also, it is customary to account for the $(1+z_\oplus)$ factor at the stage of wavelength calibration in the reductions, or in the fitting of spectral features, so that the measured redshifts are really defined by

$$(1 + z_i) = (1 + \bar{z})(1 + z_{{\rm p},i}). \tag{5.3}$$

Expanding the right-hand side and subtracting 1 on each side gives

$$z_i = \bar{z} + (1 + \bar{z})z_{{\rm p},i} \tag{5.4}$$

The peculiar velocities are such that their average over the galaxies of the cluster is zero: $\langle z_{{\rm p},i} \rangle = 0$. Therefore, taking the average of both sides of Equation 5.4 yields

$$\bar{z} = \langle z_i \rangle \,. \tag{5.5}$$

In other words, the cosmological redshift is the average redshift of the member galaxies in the cluster.

The cluster velocity dispersion that we wish to measure is the dispersion of the proper velocities $v_{\mathrm{p},i} \equiv cz_{\mathrm{p},i}$. Solving Equation 5.4 for $z_{\mathrm{p},i}$ gives

$$v_{\mathrm{p},i} = cz_{\mathrm{p},i} = \frac{c(z_i - \bar{z})}{1 + \bar{z}}. \tag{5.6}$$

In other words, the redshift differences that we measure need to be corrected for the cosmological redshift to yield correct proper velocities.

5.1.2 Dispersion and velocity errors

Measurement errors in the individual galaxy redshifts potentially affect the measurement of the dispersion. This effect is explained in detail in Danese et al. (1980), and the associated correction has become a standard step in the calculation of the velocity dispersion. For the spectral resolution of our observations, this error correction is not significant, and will correct a measured dispersion of 1000 $\mathrm{km\,s^{-1}}$ by at most several $\mathrm{km\,s^{-1}}$. We present it here nonetheless, for two reasons. First, its calculation was part of the analysis code. Second, the author has also participated in non-SPT study (Brodwin et al., 2011) with lower resolution *Keck*/LRIS and *Hubble*/WFC3 grism spectroscopy where this correction was important.

The effect is statistically well-known: random errors do not affect the calculation of averages, but enlarge variances. If we think of the distribution of peculiar velocities as a normal distribution and of the measurement errors as a smaller normal distribution, then the distribution of observed velocities will be the convolution of the two, which has the same mean but a larger variance than the "true" variance.

Let us write the redshifts as the true redshift plus an error:

$$z_i = z_{i,\mathrm{true}} + \Delta z_i. \tag{5.7}$$

Then

$$\bar{z} = \langle z_i \rangle = \langle z_{i,\mathrm{true}} \rangle + \langle \Delta z_i \rangle \,. \tag{5.8}$$

On average, $\langle \Delta z_i \rangle = 0$, so that the cosmological redshift is unbiased by errors, as would be expected.

The measured peculiar velocity is

$$v_{\mathrm{p},i} = \frac{c(z_{i,\mathrm{true}} + \Delta z_i - \langle z_j \rangle)}{1 + \langle z_j \rangle} \tag{5.9}$$

The measured variance is

$$\begin{aligned}
\sigma^2 &= \frac{1}{N-1} \sum_{i=1}^{N} v_{\mathrm{p},i}^2 && (5.10)\\
&= \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{c(z_{i,\mathrm{true}} + \Delta z_i - \langle z_{j,\mathrm{true}} \rangle - \langle \Delta z_j \rangle)}{1 + \langle z_{j,\mathrm{true}} \rangle + \langle \Delta z_j \rangle} \right)^2 && (5.11)\\
&= \frac{1}{N-1} c^2 \left[\frac{1}{1 + \langle z_{j,\mathrm{true}} \rangle} + O\left(\frac{\langle \Delta z_j \rangle}{(1 + \langle z_{j,\mathrm{true}} \rangle)^2} \right) \right]^2 \\
&\quad \times \sum_{i=1}^{N} \left[\left(z_{i,\mathrm{true}} - \langle z_{j,\mathrm{true}} \rangle \right)^2 + \left(\Delta z_i - \langle \Delta z_j \rangle \right)^2 \right] && (5.12)\\
&\simeq \sigma_{\mathrm{true}}^2 + \frac{c^2}{(1+\bar{z})^2} \frac{1}{N-1} \sum_{i=1}^{N} (\Delta z_i - \langle \Delta z_j \rangle)^2 && (5.13)\\
&= \sigma_{\mathrm{true}}^2 + \frac{c^2 \left\langle (\Delta z_i)^2 \right\rangle}{(1+\bar{z})^2} && (5.14)
\end{aligned}$$

In Equation 5.12, the denominator of the previous line was expanded as a Taylor series, and the cross-term in the expansion of the square inside the summation was omitted because as a first-order term in $\langle \Delta z_i \rangle$, it is ultimately vanishing. In the last

line, the sample variance of the Δz_i was replaced by their population variance, with known population mean of zero.

Therefore, the amount of $c^2 \langle (\Delta z_i)^2 \rangle / (1 + \bar{z})^2$ needs to be subtrated from the measured σ^2.

5.1.3 Phase-space membership selection

As explained in Chapter 1, the bulk dynamical properties of halos, such as the velocity dispersion and its scaling relation with mass, have been investigated in N-body simulations, where the dispersion can be defined in a precise physical radius of the cluster.

In real data, we need a prescription to separate the galaxies that are in the cluster from the galaxies that aren't, based on their observed phase-space coordinates. This step is usually called membership selection, although it might more aptly be named something else, for instance membership classification, as it applies to the analysis of the observed data.

Interlopers

One important thing to appreciate and understand about membership selection is the presence of *interlopers*, galaxies that occupy the same projected phase-space location as the cluster, but are not physically members of the cluster.

Figure 5.1, taken from Mamon et al. (2010), shows the line-of-sight velocity ($v_{\rm los}$) of galaxies as a function of the line-of-sight distance ($D_{\rm los}$) in a simulated light-cone that contains a cluster of galaxies, the concentration of galaxies around $(0, 0)$. Away

from the cluster, the galaxies follow the Hubble Law, $v_{\rm los} \propto D_{\rm los}$. The cluster galaxies have a wide range of velocities, corresponding to the velocity dispersion of the cluster. The figure shows that a velocity cut, for instance the 3σ cut shown as dashed red lines in Figure 5.1, cannot separate all field galaxies from cluster galaxies and will retain field galaxies whose Hubble velocity is consistent with large velocities within the cluster; these are the *interlopers*. In real data, $D_{\rm los}$ is not known, so the interlopers and cluster galaxies are not separable.

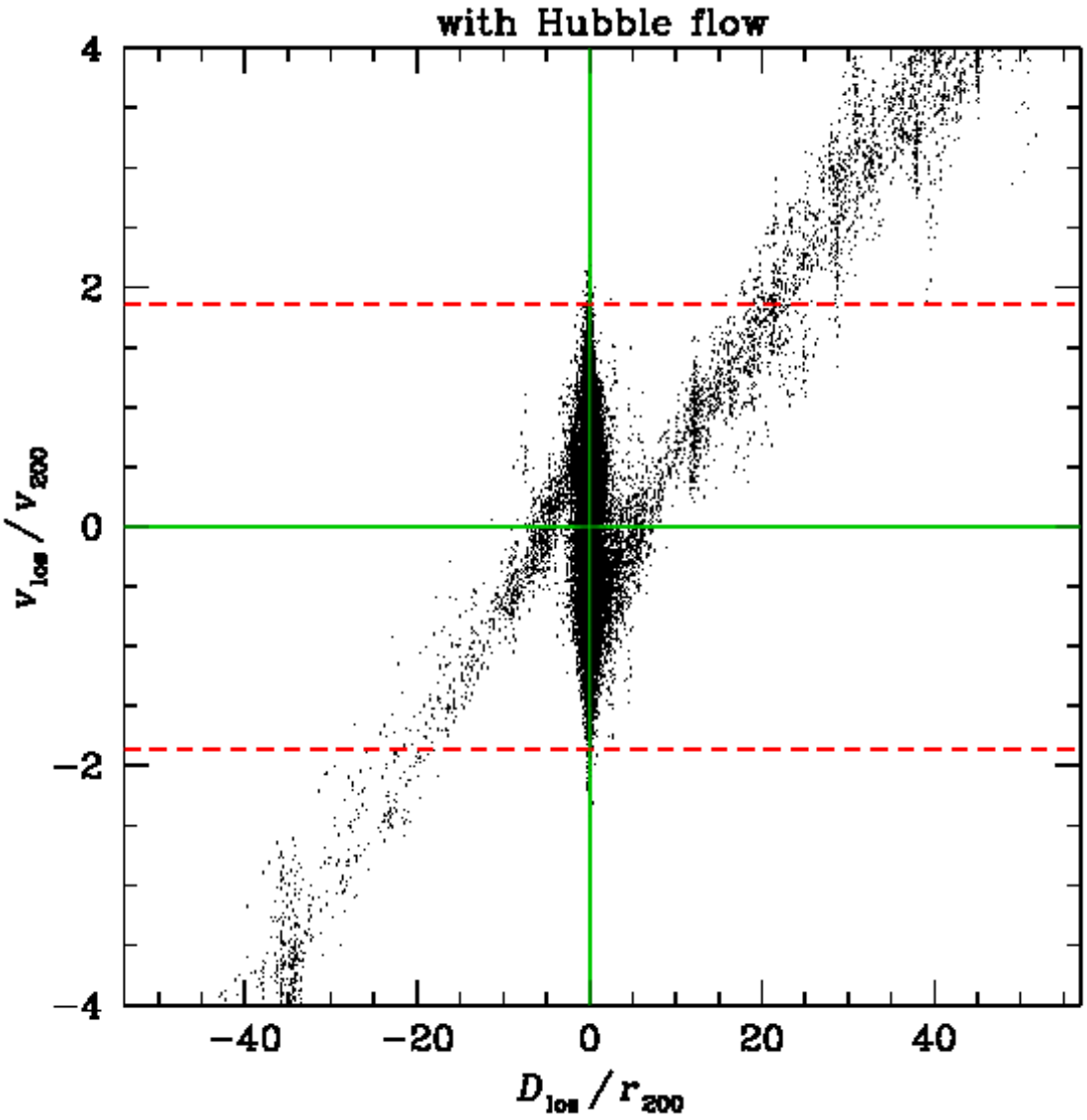

Figure 5.1: This figure from Mamon et al. (2010) shows the line-of-sight velocity of galaxies (as would be observable from spectroscopy) as a function of the line-of-sight distance (not observable, except for the galaxies closest to the Earth) in a simulated light-cone that contains a cluster of galaxies. The axes of velocity and distance are centered on the cluster. The red dashed lines enclose the galaxies that would be selected as cluster members by a three-sigma cut in proper velocity.
Image credit: Mamon et al. (2010).

Choice of membership selection algorithm

There are many prescriptions in the literature for separating cluster members from non-members, in the form of a velocity cut for all galaxies, or a (projected-)radius-dependent velocity cut. Many of them are physically motivated and based on years of experience in the community, and give similar results on average (Girardi et al., 1993; White et al., 2010), but ultimately we would like our choice of membership algorithm to be precisely characterized by simulations to make sure that this step does not introduce a bias.

Two algorithms are tested from recent simulations in the literature. The first one is a radially-dependent velocity cut (den Hartog & Katgert, 1996; Biviano et al., 2006; White et al., 2010), which we call "the phase-space method" in accordance with the last reference. It essentially relies on using the observed velocities and positions of the galaxies to construct a velocity dispersion *profile*, and use that profile to construct a velocity cut dependent on the projected radius from the center of the cluster. In terms of the velocity dispersion, it is close to a 2σ cut, and it is shown for our stacked cluster in Figure 5.2. Reliance on the profile makes it not well fit to our program, given our small sample sizes. We note that White et al. (2010) successfully apply it to samples as small as 25 members. Another consideration is that these studies (like most studies from simulations) sample observations from dark matter particles (or better, dark-matter subhalos in White et al., 2010), so that the non-member and interloper fractions and radial properties are likely to be different from our observations, where we have observed mostly bright red-sequence galaxies.

Application of the phase-space method to our data yields a reasonable selection

for most clusters, but also leads to a number of catastrophic failures where most of the data points are rejected, and where the final dispersion is unphysically low, around 300 $\mathrm{km\,s^{-1}}$. This includes the cluster for which we have the most redshifts, SPT-CL J2331-5051. We conjecture that the algorithm would need to be changed for a red-sequence selection, but these failures my also be due to cluster substructure.

The second algorithm that is tested in simulations and that is the natural choice for our observations is 3-sigma clipping (Yahil & Vidal, 1977; Mamon et al., 2010; Saro et al., 2012). Saro et al. (2012) employ mock red-sequence observations from simulations and explore the interloper properties after 3σ clipping has been applied. The interloper fraction is smaller than for random dark matter selection (see, e.g., Mamon et al., 2010). The measured velocity dispersion is biased high, and depending on parameters such as the aperture of the observations, the mass and redshift of the cluster, and magnitudes of the spectroscopic galaxies, the velocity bias ranges from minus a few percent to about 10%. This may seem counter-intuitive, as the red galaxies have had more time to be affected by dynamical friction in the cluster, and indeed red-sequence member galaxies have a smaller velocity dispersion on average than blue member galaxies (see, e.g., Girardi et al., 1996; Fadda et al., 1996; Hwang & Lee, 2008). However it is the combination of spatial selection, color selection and membership selection that will determine how the measured dispersion relates to the true dispersion for the halo.

In Figure 5.1, a 3σ cut in galaxy-cluster velocities is shown by red dashed lines; it is also a black dashed line in the plots showing the stacked cluster, Figures 5.2 and 5.3.

5.1.4 Resistant and robust estimators

Velocity studies of clusters of galaxies are inherently difficult in that they are dealing with ill-characterized systematic uncertainties in the sampling, membership and source velocity distribution. As a result, it has become customary to calculate the average (cosmological) redshift and the velocity dispersion in ways that are, as much as possible, insensitive to those effects.

Beers et al. (1990) describes in some detail the properties of *resistant* and *robust* estimators. Resistance means that the estimate does not change much when a number of data points are replaced by other values; the median is a well-known example of a resistant estimator. Robustness means that the estimate does not change much when the distribution from which the data points are drawn is varied.

The cluster redshift is usually calcluated via the biweight average, which is given by

$$z_{\rm BI} = M + \frac{\sum_{|u_i|<1}(1-u_i^2)^2(z_i - M)}{\sum_{|u_i|<1}(1-u_i^2)^2} \tag{5.15}$$

with

$$u_i = \frac{z_i - M}{6{\rm MAD}(z_i)}, \tag{5.16}$$

where M is the median redshift, and ${\rm MAD}(z_i)$ is the median absolute deviation of the redshifts:

$${\rm MAD}(z_i) = {\rm median}(|z_i - M|). \tag{5.17}$$

For normally distributed random variables, $\sigma \simeq 1.48{\rm MAD}$, so that $6{\rm MAD} \simeq 4\sigma$.

The biweight average is part of a mathematical family of location estimators called M-estimators. With a proper choice of weighting functions, M-estimators can be constructed that interpolate smoothly between the median and the average, and

the biweight retains similarities to both. As can be seen from the formula above, it is really a weighted average with a weighting function that does a hard sigma-clipping around the median, with sigma estimated through the median absolute deviation. The weighting function $(1-u_i^2)^2$ was called the "bi-square weight" because of the two squares, hence the name *biweight*.

Beers et al. (1990) also presents the formula for the biweight velocity dispersion. However, that estimator (or more correctly, the associated variance) is biased for samples, in the same way that calculating the variance of data points x_i as $\frac{1}{N}\sum(x_i - \langle x_i \rangle)^2$ would yield a biased value for a sample, compared to the true value σ^2, and the *sample variance* $\frac{1}{N-1}\sum(x_i - \langle x_i \rangle)^2$ is preferred in that case to yield an unbiased estimate of the underlying population's variance.

Anecdotally, we know that the fact that the Beers et al. (1990) estimator is biased is often recognized by researchers who use an unbiased version and nonetheless simply quote Beers et al. (1990), making it in effect impossible to know for sure which version any study from the literature has used. Also, the implementation of the Fortran code companion to Beers et al. (1990)[1] contains a partial correction of this bias, in a factor of $\sqrt{n/(n-1)}$ multiplying the dispersion.

We use the biweight sample variance (see, e.g., Mosteller & Tukey, 1977)

$$\sigma_{\rm BI}^2 = N_{\rm members}\frac{\sum_{|u_i|<1}(1-u_i^2)^4(v_i-\bar{v})^2}{D(D-1)} \tag{5.18}$$

where v_i are the proper velocities, $\bar{v}$ their average,

$$D = \sum_{|u_i|<1}(1-u_i^2)(1-5u_i^2), \tag{5.19}$$

[1] rostat.f, version 1.2, February 1991. Retrieved April 2012 from http://www.pa.msu.edu/ftp/pub/beers/posts/rostat/rostat.f

and u_i is the biweight weighting

$$u_i = \frac{v_i - \bar{v}}{9\mathrm{MAD}(v_i)}. \tag{5.20}$$

Because it comes at no cost in computation or complexity, we argue that this biweight sample variance is the biweight estimator that should always be used. It is perhaps a bit of a pedantic point for those studies that have large numbers of redshifts, but this statistical bias is certainly an effect that we wish to avoid in the few-N_{members} regime.

5.2 The stacked cluster

We produce a stacked cluster from our observations, as a way to look at the ensemble phase-space galaxy selection; this stacked cluster will also be useful for resampling tests, Section 5.3. We generate it in a way that is independent of cluster membership determination. As membership selection and the calculation of the velocity dispersion are unavoidably intertwined, we use the SZ- and X-ray-based SPT mass, the other uniform mass measurement that we have for all clusters, to normalize the velocities before stacking. This eliminates any effect that interlopers or dispersion errors due to sampling bias would have in the stacking.

We make a stacked proper-velocity distribution independent of any measurement of the velocity dispersion by calculating the "equivalent dispersion" from the SPT mass. We convert the $M_{500c,\mathrm{SZ}}$ to $M_{200c,\mathrm{SZ}}$ assuming an NFW profile and the Duffy et al. (2008) concentration, and then convert the $M_{200c,\mathrm{SZ}}$ to a σ_{SZ} (in $\mathrm{km\,s^{-1}}$) using the Saro et al. (2012) scaling relation. This σ_{SZ} is listed for each cluster in Table

6.2 for reference. We also normalize the distance to the SZ center by $R_{200c,\mathrm{SZ}}$. The resulting phase-space diagram of the normalized proper velocities v_i/σ_{SZ} vs $r_i/R_{200c,\mathrm{SZ}}$ is shown in Figure 5.2. For reference, different velocity cuts are plotted. The black dashed line is a 3-sigma cut. The red dot-dashed line is the "phase-space method" (den Hartog & Katgert, 1996; Biviano et al., 2006; White et al., 2010) velocity cut, where the mass as a function of radius is calculated from an NFW profile of typical concentration instead of the dynamical mass; it is close to a 2σ cut. The blue dotted line is a radially-dependent $2.7\sigma(R)$ cut, where again the $\sigma(R)$ is from an NFW profile; this velocity cut is found to be optimal for rejecting interlopers by Mamon et al. (2010) (although when considering systems without red-sequence selection). All of these cuts would be applied iteratively in membership selection.

The histogram of proper velocities is shown in the right panel, together with a Gaussian of mean zero and standard deviation of one. This normal curve is the expected distribution for randomly chosen cluster members, if σ_{SZ} is close to the true velocity dispersion, on average. The similar shape and the low number of data points with a large proper velocity is an encouraging sign of the success of the selection.

Figure 5.3 shows the stacked cluster using the proper velocities normalized by the dispersion measured after 3σ membership selection, and the points and crosses show the 3σ members and non-members, respectively. Interestingly, the phase-space method cut seems to delineate the main envelope of the members quite well at projected radius $\gtrsim 0.4R_{200c}$, but not at smaller radius.

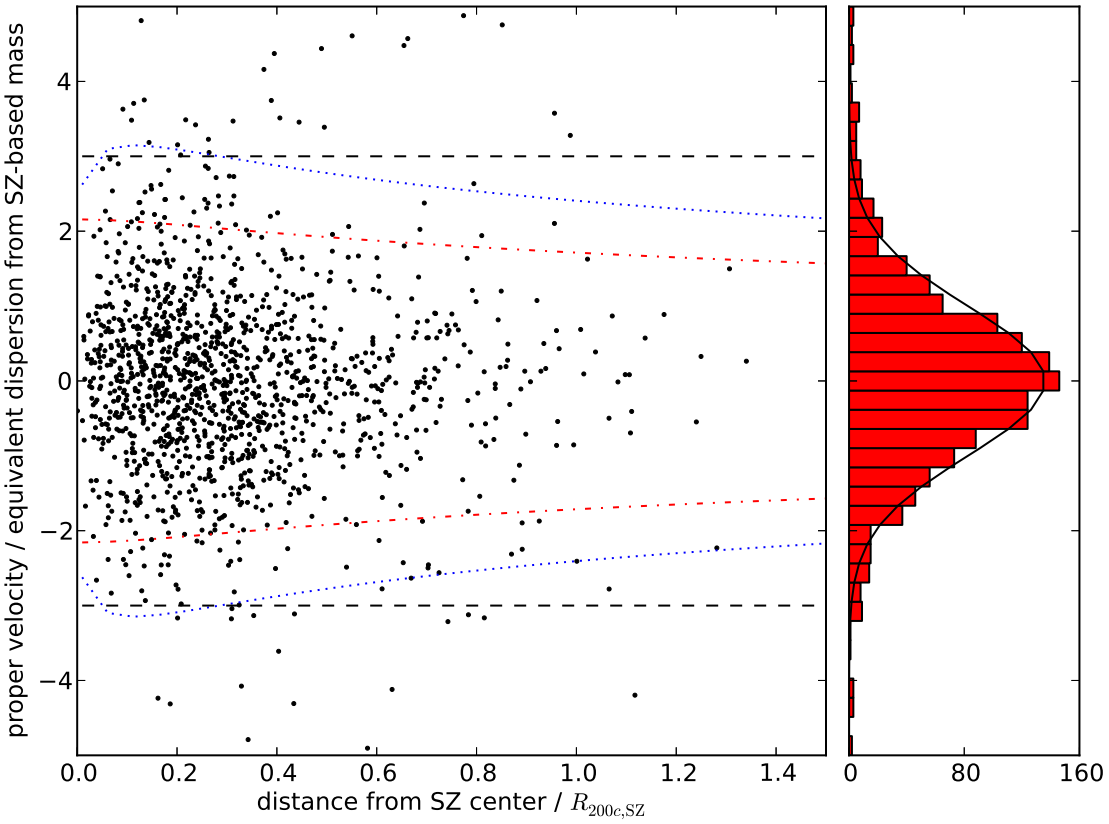

Figure 5.2: Stacked cluster, using the dispersion equivalent to the SPT mass. *Left panel:* phase-space diagram of velocities. The red dot-dashed line is "the phase-space method" velocity cut, the blue dotted line is a radially-dependent $2.7\sigma(R)$ cut, and the black dashed line is a 3-sigma cut. All of these cuts would be applied iteratively in membership selection. *Right panel:* histogram of proper velocities, with a Gaussian distribution with a mean of zero and standard deviation of one, and area equal to that of the histogram.

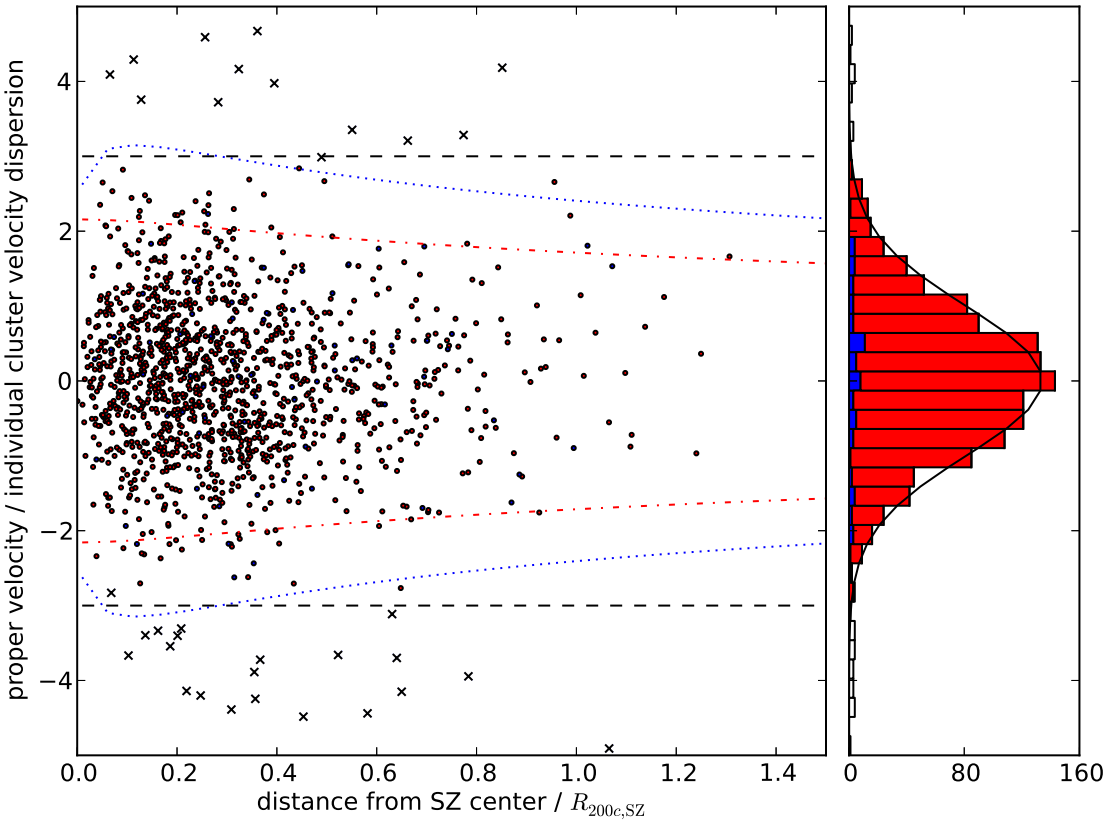

Figure 5.3: Stacked cluster, using the dispersion measured in each individual cluster. *Left panel:* phase-space diagram of velocities. Non-members are shown as crosses, passive galaxies as red dots, and emission-line galaxies as blue dots. The red dot-dashed line is "the phase-space method" velocity cut, the blue dotted line is a radially-dependent $2.7\sigma(R)$ cut, and the black dashed line is a 3-sigma cut. All of these cuts would be applied iteratively in membership selection. *Right panel:* histogram of proper velocities, with a Gaussian distribution with a mean of zero and standard deviation of one, and area equal to that of the histogram. Red and blue are the galaxies with continuum and emission-line redshifts, respectively.

5.3 Statistical methodology in the few-$N_{\rm members}$ regime

In this section, we explore the statistical issues surrounding our obtaining reliable estimates of velocity dispersions and associated confidence intervals. Our results are not new knowledge as the science of statistics is concerned, but we sort through a few issues that are unclear or inconsistent in the velocity dispersion literature and can become important in the few-$N_{\rm member}$ regime, where, for instance, we need to be attentive to the fact that N and $(N-1)$ differ by enough to perturb our results. A key element in our approach is "resampling", in which we extract and analyze subsets of the data, either on a cluster-by-cluster basis, or from the stacked cluster that we constructed from the entire catalog. This allows us to generate large numbers of "pseudo-observations" to address statistical questions where we have too few observations to directly answer.

5.3.1 Generated subsamples

We generate sub-samples with different $N_{\rm members}$ by resampling from two different source distributions.

First, we use the individual clusters for which we obtained 30 or more member velocities as source distributions from which we randomly extract smaller samples, as though we had observed fewer member galaxy redshifts. The cluster redshift and dispersion from those smaller, random samples can be computed and compared to the value that was measured with the full data set. This reference value is not the true

dispersion of the halo and itself carries an uncertainty. Specifically, for each of the 17 clusters for which we obtained 30 or more member velocities, we randomly draw 500 "pseudo-observations" with $8 \leq N_{\rm members} \leq 28$. These subsamples are different but not independent, which we take into account in our analysis.

Second, from the stacked cluster (with a velocity cut of 5σ), we resample a large number of pseudo-observations with $8 \leq N_{\rm members} \leq 30$, on which we apply an iterative 3σ membership selection cut.

Resampling from the individual clusters as opposed to the stack preserves possible cluster substructure that would be statistically diluted in the stack, and also offers our best membership selection. Resampling from the stack is the best way that we have of generating different observations of a same cluster, which is the average cluster.

5.3.2 Unbiased estimators

Estimators and confidence intervals for velocity dispersions are discussed in Section 5.1.4. As we are exploring the properties of the few-$N_{\rm members}$ regime, we would like our estimators to be *unbiased*, in addition to being robust and resistant, meaning that the mean estimate should be independent of the number of points that are sampled. In other words, limiting observations to a small number of members per cluster should not introduce biases, beyond yielding larger statistical uncertainties.

We discussed part of this issue already in Section, 5.1.4, when noting that the expression for the biweight variance from Beers et al. (1990) is biased for samples, and that the biweight sample variance, equation 5.18, is preferable, much like the "regular", Gaussian case where there are separate estimators that are the population

variance and the sample variance.

But statistics are subtle and bias can come in different ways. The biweight sample variance (like the regular sample variance) is an unbiased estimator of variance but not of the dispersion (standard deviation). In other words, when sampling a Gaussian distribution of variance σ^2, the following expectations values hold:

$$\mathrm{E}(\sigma_{\mathrm{BI}}^2) = \sigma^2 \text{ but } \mathrm{E}(\sigma_{\mathrm{BI}}) \neq \sigma. \tag{5.21}$$

This statement is N_{members}-dependent, and for any reasonable N_{members} (such as more than 15), effectively $\mathrm{E}(\sigma_{\mathrm{BI}}) = \sigma$. It is nonetheless for this reason that we calculate and quote (in Table 6.2) confidence intervals from resampling that are symmetric around σ_{BI}^2, as this is the natural space of the estimator, where the resampled distribution is closest to being symmetric at all N_{members} (see Section 5.3.3, and Figures 5.7 and 5.8). The gapper scale estimator (Beers et al., 1990) is an unbiased estimator of dispersion – not variance –, so we present confidence intervals on σ_{G}, not σ_{G}^2.

The solid black line in Figures 5.4 and 5.5 is the average value of biweight estimates as a function of N_{members} for our resampled pseudo-observations (we will discuss the other lines, error and uncertainty, in Section 5.3.3). Figure 5.4 shows the biweight average of the galaxy redshifts, normalized by the cosmological redshift of the cluster, their "asymptotic value", for the pseudo-observations resampled from the stacked cluster, where a 3σ membership selection cut is applied to every subsample. The average shows no dependence on N_{members}. Generating the same figure from the pseudo-observations resampled from individual clusters looks nearly identical, which shows that cluster substructure and membership selection have no adverse effect on the determination of the cosmological redshift.

Figure 5.5 shows the (normalized) biweight sample variance, this time for the pseudo-observations resampled from the individual observed clusters[2]. The average departs from unity at small $N_{\rm members}$ for some of the clusters where the value of the dispersion can show a tendency to be high or low by a few percent, but on average the line is very close to unity.

[2]Figure 5.6 is the equivalent figure from stacked-cluster resampling, which we will discuss when talking about the effect of membership selection.

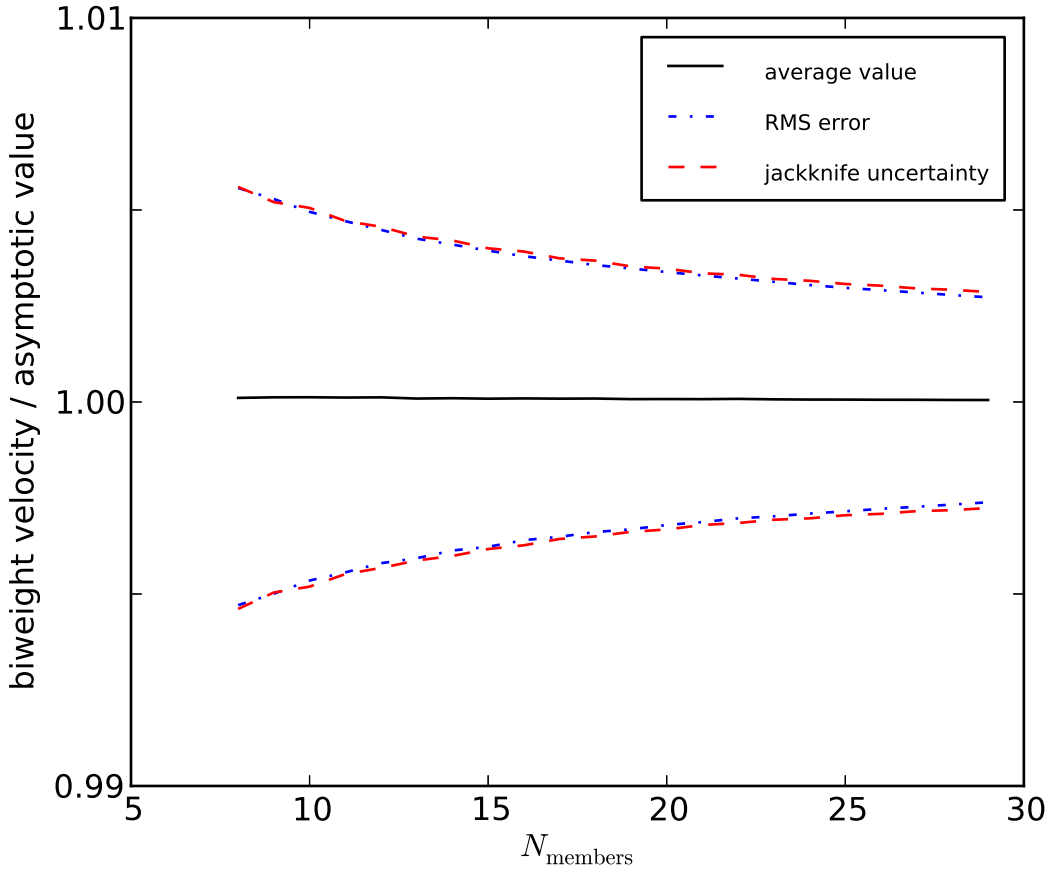

Figure 5.4: Biweight average of the galaxy redshifts, normalized by the cosmological redshift of the cluster, their "asymptotic value", for the pseudo-observations resampled from the stacked cluster, where a 3σ membership selection cut is applied to every subsample. The solid black line is the average value, the dashed red line is the average jackknife confidence interval, and the dot-dashed blue line is the sum of the error from the asymptotic value plus the purely Gaussian resampling uncertainty in the asymptotic value. The average is independent of N_{members}. Generating the same figure from the pseudo-observations resampled from individual clusters looks nearly identical, showing that cluster substructure and membership selection have no adverse effect on the determination of the cosmological redshift.

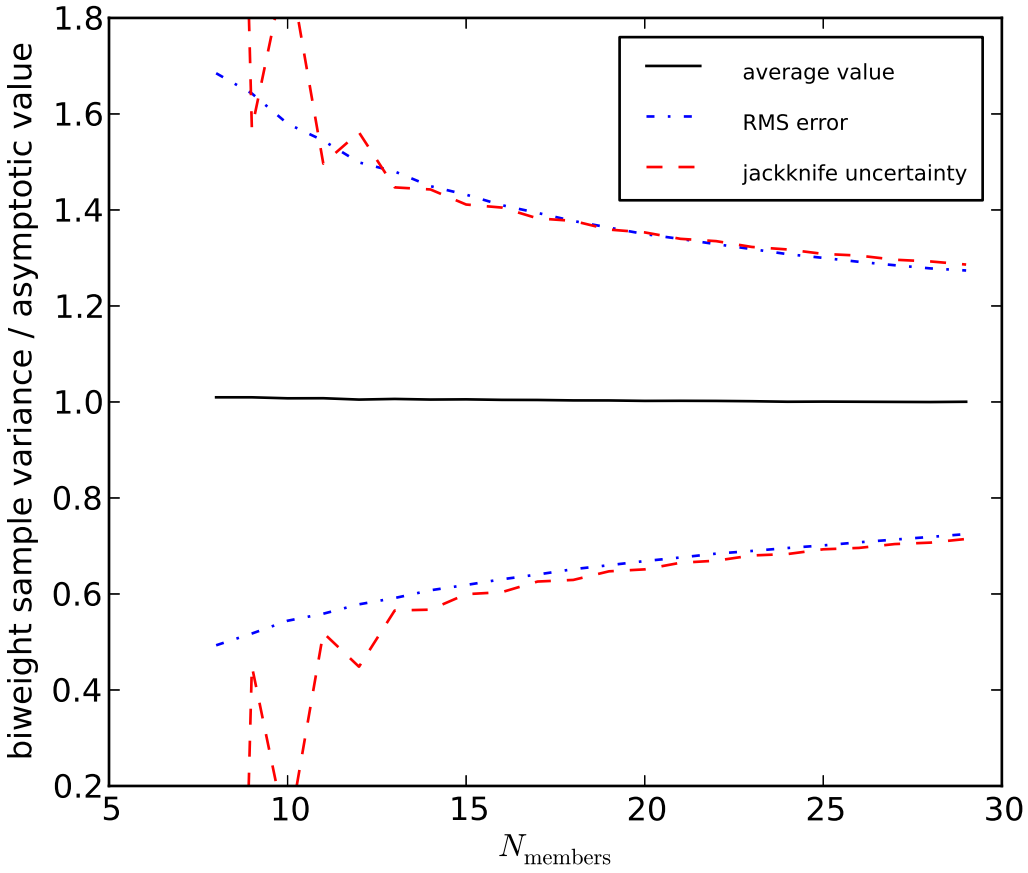

Figure 5.5: Biweight sample variance normalized by the asymptotic value (that is, the measured value reported in Table 6.2), and averaged over an ensemble of resampled clusters. There are 500 resampled clusters at each N_{members} for each of the 17 clusters with 30 or more members. The solid black line is the average value, the dashed red line is the average jackknife confidence interval, and the dot-dashed blue line is the sum of the error from the asymptotic value plus the purely Gaussian resampling uncertainty in the asymptotic value. The average departs from unity at low N_{members} for some of the clusters where the value of the dispersion, in particular, can show a tendency to be high or low by a few percent, but on average the line is very close to unity.

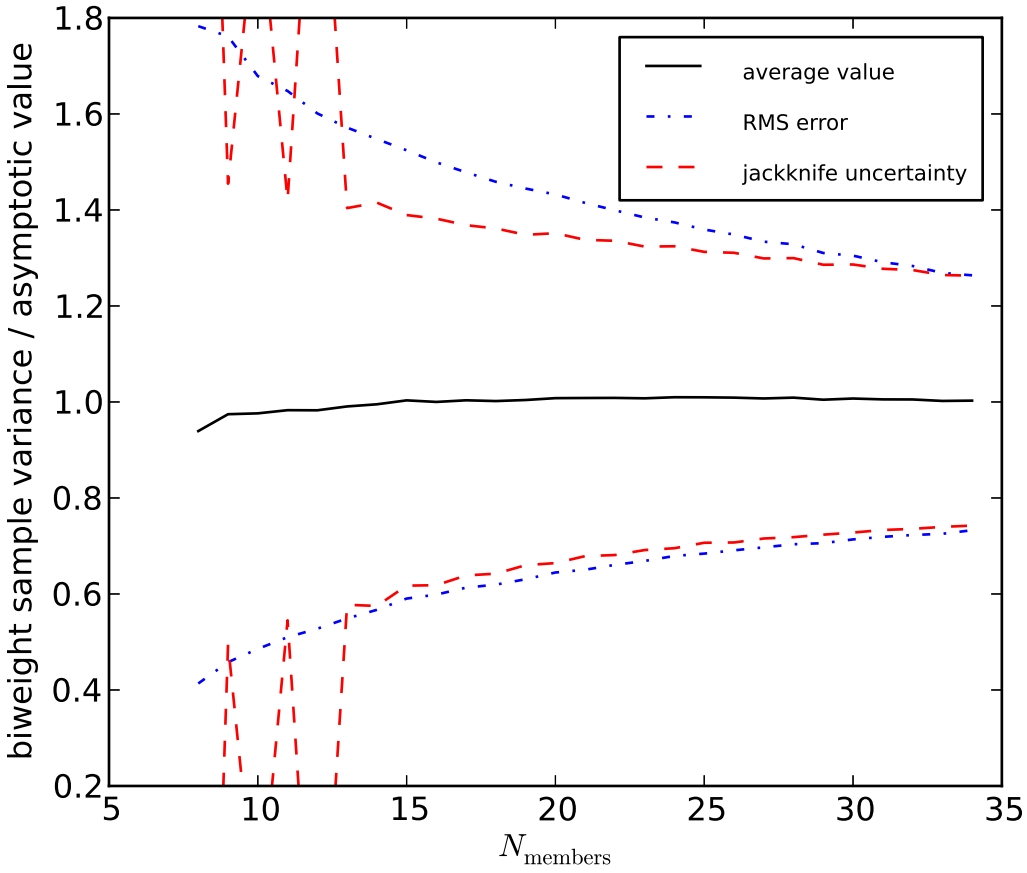

Figure 5.6: Biweight sample variance normalized by the asymptotic value, and averaged over an ensemble of pseudo-observations resampled from the stacked cluster, where a 3σ membership selection cut is applied to every subsample. The solid black line is the average value, the dashed red line is the average jackknife confidence interval, and the dot-dashed blue line is the sum of the error from the asymptotic value plus the purely Gaussian resampling uncertainty in the asymptotic value.

5.3.3 Confidence intervals

In addition to introducing biweight estimators, Beers et al. (1990) describe a number of different ways in which the confidence intervals on biweight estimators can be calculated; most notably, they give a good, short introduction to the statistical jackknife and the statistical bootstrap.

The *statistical jackknife* (see, e.g. Mosteller & Tukey, 1977) constructs a confidence interval for an estimate from how much it varies when data points are removed. The "delete-1" jackknife algorithm, which we just call "statistical jackknife" for simplicity, is the following. For an estimator f (e.g. a flavor of variance, $\sigma^2_{\rm BI}(v)$) and N data points v_i, generate N *pseudovalues* $y^\star_i = f(v_{(i)})$ where $v_{(i)}$ is the dataset with the ith value removed. Then the confidence interval is

$$\sigma_f = \frac{1}{N(N-1)} \left(\sum (y^\star_i)^2 - \frac{1}{N} (\sum y^\star_i)^2 \right). \tag{5.22}$$

The *statistical bootstrap* generates a probability distribution function for the estimate from resampling the observed values with replacement a large number of times, often 1000 or more. For instance, from 30 observed galaxy redshifts, one could generate the p.d.f. of the cosmological redshift and velocity dispersion by generating 1000 samples of 30 randomly chosen galaxy redshifts, where there are possibly duplicate data points in each sample, and calculate the cosmological redshift and velocity dispersion for each sample. The confidence intervals can then be found from the percentiles of this distribution. This would seem like an intuitive procedure, but it is not always so simple. Bootstrap distributions often exhibit bias; this bias is well understood by statisticians who have developped prescriptions for correcting it (see, e.g., Efron, 1987). This bias does not affect the biweight average, but in the case of

the biweight dispersion, the median of the statistical bootstrap distribution is biased with respect to the value of the estimate, which often leads to strongly asymmetric confidence intervals, even with second-order ("BCa") bias correction (Efron, 1987). It was not immediately clear to us whether those confidence intervals are meaningfully asymmetric, or whether the asymmetry is just an artifact left from incomplete bias subtraction. It could be meaningful: because the dispersion cannot be negative (the same is true of the variance), the distribution needs to be asymmetric to some extent.

Many publications following Beers et al. (1990) have chosen to calculate confidence intervals with the statistical bootstrap, and different practices can be seen in its use. When applied to the dispersion, Fadda et al. (1996) and Girardi et al. (1996) for instance, quote asymmetric intervals where the upper interval is larger than the lower interval, as in 1053^{+164}_{-108} km s^{-1}. Others quote symmetric intervals from the bootstrap (Zhang et al., 2011; Sifón et al., 2012), as in (1053 ± 139) km s^{-1}, which mathematically means that they ignore the actual value of the percentiles in the bootsrap distribution, and only use its width symmetrically around their estimate. This width is usually similar to the size of the confidence interval obtained from the statistical jackknife, which is symmetric.

To complicate things further, if the confidence intervals are calculated symmetrically for $\sigma^2_{\rm BI}$ instead of $\sigma_{\rm BI}$, which makes intuitive sense given the fact explained earlier that the variance, and not the dispersion, is the unbiased estimator, then converting to a confidence interval on $\sigma_{\rm BI}$ creates an asymmetry where the lower interval is larger than the upper interval, for instance 1053^{+108}_{-164} km s^{-1}, which is the opposite of what the bootstrap creates.

We use resampling to determine which procedure is the best, that is, which one produces confidence intervals that most accurately reflect the statistical variation in the velocity dispersion measurement.

We use the samples drawn from the stacked cluster to explore the distribution of the variance ($\sigma^2_{\rm BI}$). Figure 5.7 shows the histogram of 1000 $\sigma^2_{\rm BI}$ measurements with $N_{\rm members} = 30$. The jackknife (red dashed), bootstrap (green dash-dotted) and jackknife from the log variance (not shown[3]) all have a similar size to the RMS error. The jackknife does not capture the slight asymmetry, while the bootstrap overestimates the asymmetry.

We do the same calculations and plot for the logarithm of the variance ($\ln \sigma^2_{\rm BI} \propto \ln \sigma_{\rm BI}$), and the results are shown in Figure 5.8. It is a natural thing to look at, because scaling relation calculations (and generally speaking, most plots) involve the logarithm of the dispersion.

The distribution has a low tail and is not as symmetric as the distribution of the variance, althought the estimate of symmetry is not so bad at the one-sigma level.

In view of this, we choose to quote confidence intervals in $\sigma^2_{\rm BI}$. We choose the statistical jackknife; the statistical bootstrap appears to be performing similarly, but it is computationally much more expensive.

We now wish to verify that the jackknife offers a good way to determine confidence intervals down to few $N_{\rm members}$. In Figures 5.4 and 5.5, the root-mean-square error of the estimator from the asymptotic value is shown as a dash-dotted line. Because of

[3]As σ^2 cannot be negative, Beers et al. (1990) suggest that the log variance is a better space to calculate confidence intervals. We do not find that is is the case from the shape of the distribution, but we note that all of our clusters are clearly very massive, so we are not in a regime of marginal cluster detections with dispersion possibly consistent with zero.

the large covariance of the subsample, the curve shown as the RMS error is actually the measured RMS error between the subsamples and a symmetric the $N_{\rm members} = 30$ uncertainty (from Figure 5.7), added in quadrature.

Then, for each of the subsamples, the uncertainty was computed using the statistical jackknife. The average jackknife uncertainty is shown as a red dashed line. It follows the RMS error well. In the case of the biweight sample variance, the jackknife becomes less reliable for $N_{\rm members} \simeq 10 - 12$ and less.

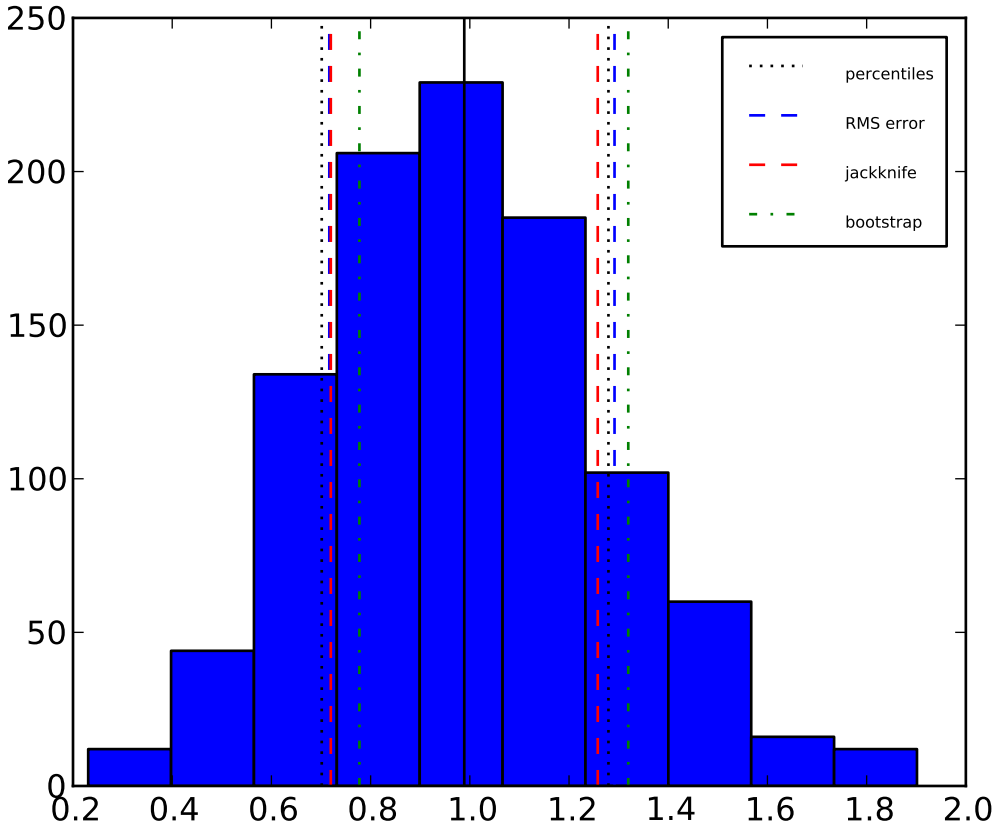

Figure 5.7: Distribution of 1000 measurements of the velocity variance ($\sigma^2_{\rm BI}$) with $N_{\rm members} = 30$ galaxy redshifts randomly sampled from the stacked cluster. This distribution is more symmetric than the distribution of $\sigma_{\rm BI}$, or $\ln \sigma_{\rm BI}$ (Figure 5.8). The values of $\sigma^2_{\rm BI}$ are normalized by the variance of the velocities of the entire stacked cluster. The solid line shows the mean which is very close to 1. The dotted line shows the 68% confidence interval as computed from percentiles of the distribution. The blue dashed line is the asymmetric root-mean-square error of the subsample measurements. For each subsample, the jackknife and boostrap (with 1000 iterations) confidence intervals were computed. Their average is shown as a red dashed line for the jackknife, and an green dash-dotted line for the bootstrap.

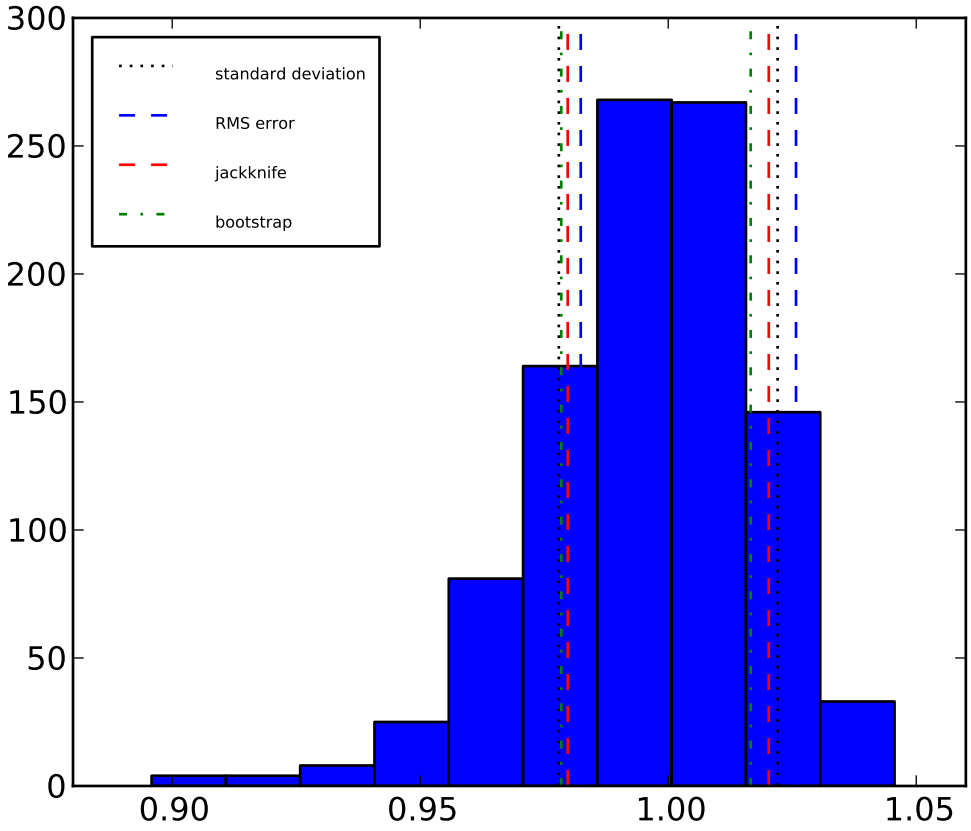

Figure 5.8: Distribution of 1000 measurements of the logarithm of the velocity variance ($\ln \sigma^2_{\rm BI} \propto \ln \sigma_{\rm BI}$) with $N_{\rm members} = 30$ galaxy redshifts randomly sampled from the stacked cluster. This distribution is to be compared with the distribution of $\sigma^2_{\rm BI}$ (Figure 5.7). The values of $\sigma^2_{\rm BI}$ are normalized by the log of the variance of the velocities of the entire stacked cluster. The dotted line shows one sigma as calculated using the standard deviation. The blue dashed line is the asymmetric root-mean-square error of the subsample measurements. For each subsample, the jackknife and boostrap (with 1000 iterations) confidence intervals were computed. Their average is shown as a red dashed line for the jackknife, and an green dash-dotted line for the bootstrap.

5.3.4 Impact of membership selection

Membership selection has the potential of changing the average of our variance measurements, as well as the shape of their distribution, and our estimates of the confidence interval.

Figure 5.6 is similar to Figure 5.5, except that the samples were drawn from the stacked cluster, and membership selection by 3σ clipping is applied before computing the variance for each $N_{\rm members}$. The RMS error is visibly asymmetric and the statistical jaccknife underestimates it. This is intuitively right as the sigma clipping will only remove variability from the data, and therefore make estimates that rely on variability in the data, like the jackknife or bootstrap, lower on average.

The mean is underestimated by a few percent when few members are used; the size of this bias is of order 1% at $N_{\rm members} = 15$ and 5% at $N_{\rm members} = 10$.

At $N_{\rm members} = 25$, which is both the median size of our samples and the point where the statistical uncertainty equals 15%, the mean statistical (jackknife) uncertainty for our sample, the combination of statistical plus systematic uncertainty due to membership selection equals 19%. Adding the 12% intrinsic scatter in quadrature yields a 22% floor for the measured scatter, before other systematics are taken into account. This is a minimum, as the errors grow larger than the jackknife uncertainty in the $N_{\rm members} = 15 \sim 20$ range.

In conclusion, the distribution of our measured velocity dispersions does not exhibit a strong bias with varying $N_{\rm members}$, and the uncertainties derived from the statistical jackknife capture the statistical variance of the data well, but not systematic errors added by the membership selection. We cannot study those systematics

further, given the small size of our sample of velocity dispersions. Ways to address this and other systematics will be discussed in Chapter 7.

Chapter 6

Data – optical spectroscopy of 60 SPT clusters

In this chapter, we present the velocity data extracted from our spectroscopic observations, as well as derivative products, most notably velocity dispersions. We also present velocity dispersions from the literature, to produce a catalog of velocity dispersions of SPT clusters.

The velocities of SPT cluster galaxies presented here include our spectroscopic measurements of 60 massive galaxy clusters, 48 of which produce velocity dispersions calculated with more than 15 member galaxies. We have already presented several of these results elsewhere (Brodwin et al., 2010; Foley et al., 2011; Williamson et al., 2011; McDonald et al., 2012; Stalder et al., 2013; Reichardt et al., 2013), where we either reported the spectroscopic redshift, or were interested in the velocity dispersion of a single cluster. These are the data obtained through 2011 in our ongoing spectroscopy program. We also list dispersions collected from the literature, most

notably those presented by Sifón et al. (2012) who reported velocity dispersions of 16 massive clusters detected by ACT, 14 of which are also SPT detections.

6.1 Results

6.1.1 Individual galaxy redshifts

A subset of individual galaxy redshifts are listed in Table 6.1. The listed galaxies are the central galaxies (see Section 6.1.3); the full sample of redshifts for both member and non-member galaxies will be available in electronic format upon publication of the results in a research journal. For each galaxy, the table lists the SPT ID of the associated cluster, a galaxy ID, right ascension and declination, the redshift and method of redshift measurement, and notable spectral features.

6.1.2 Cluster redshifts and velocity dispersions

Table 6.2 lists the cluster redshifts and velocity dispersions measured from the galaxy velocities. The cluster redshift z is the biweight average of member galaxy redshifts with an uncertainty derived from the statistical jackknife (see Section 5.3.3; for the jackknife, see e.g. Mosteller & Tukey, 1977). Once the cluster redshift is computed, the galaxy proper velocities v_i are obtained from their redshifts z_i by $v_i = c(z_i - z)/(1 + z)$, as explained in Section 5.1. The velocity dispersion $\sigma_{\rm BI}$ is the square root of the biweight sample variance of proper velocities, the uncertainty of which is also estimated via the statistical jackknife (see Section 5.3). We also report the dispersion σ_G determined from the gapper estimator (Beers et al., 1990), which

Table 6.1: Galaxy redshifts

(*This table is continued on next page.*) Redshifts of individual galaxies. This is a partial listing. The entries listed here are the central galaxies, a subset of our observations; the central galaxy selection is explained in Section 6.1.3. For each galaxy, the table lists the SPT ID of the associated cluster, a galaxy ID, right ascension and declination, the redshift z and associated uncertainty, redshift measurement method, and notable spectral features. The labels of the "z method" column are "rvsao-xc" and "rvsao-em", respectively, for the RVSAO cross-correlation to absorption features and fit to emission lines, and "template" for an in-house template-fitting method using the SDSS DR2 templates.

Associated SPT ID	Galaxy ID	Galaxy R.A. (J2000 deg.)	Galaxy Dec. (J2000 deg.)	z	z method	Spectral features
SPT-CL J0000-5748	J000059.99-574832.7	0.2500	−57.8091	0.7007 ± 0.0002	template	[O II], Ca II H&K
SPT-CL J0037-5047	J003747.30-504718.9	9.4471	−50.7886	1.0302 ± 0.0002	template	Ca II H&K
SPT-CL J0118-5156	J011824.76-515628.6	19.6032	−51.9413	0.7021 ± 0.0004	rvsao-xc	Ca II H&K
SPT-CL J0205-5829	J020548.26-582848.4	31.4511	−58.4801	1.3218 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0205-6432	J020507.83-643226.8	31.2827	−64.5408	0.7430 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0233-5819	J023300.97-581937.0	38.2540	−58.3270	0.6600 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0234-5831	J023442.26-583124.7	38.6761	−58.5235	0.4146 ± 0.0001	rvsao-xc	[O II], Ca II H&K
SPT-CL J0240-5946	J024038.38-594548.5	40.1599	−59.7635	0.4027 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0245-5302	J024524.82-530145.3	41.3534	−53.0293	0.3028 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0254-5857	J025415.47-585710.6	43.5645	−58.9530	0.4373 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0257-5732	J025720.95-573254.0	44.3373	−57.5484	0.4329 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0317-5935	J031715.84-593529.0	49.3160	−59.5914	0.4677 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J0433-5630	J043301.03-563109.4	68.2543	−56.5193	0.6946 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0438-5419	J043817.62-541920.6	69.5734	−54.3224	0.4217 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0449-4901	J044904.03-490139.1	72.2668	−49.0275	0.7949 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0509-5342	J050921.37-534212.7	77.3390	−53.7035	0.4616 ± 0.0002	template	[O II], Ca II H&K
SPT-CL J0511-5154	J051142.95-515436.6	77.9290	−51.9102	0.6488 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0516-5430	J051637.33-543001.5	79.1556	−54.5004	0.2970 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0528-5300	J052805.29-525953.1	82.0220	−52.9981	0.7670 ± 0.0002	template	Ca II H&K
SPT-CL J0534-5937	J053430.04-593653.8	83.6252	−59.6150	0.5757 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0551-5709	J055135.58-570828.6	87.8983	−57.1413	0.4243 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J0559-5249	J055943.19-524926.2	89.9300	−52.8240	0.6104 ± 0.0002	template	Ca II H&K

Table 6.1: Galaxy redshifts (continued)

Associated SPT ID	Galaxy ID	Galaxy R.A. (J2000 deg.)	Galaxy Dec. (J2000 deg.)	z	z method	Spectral features
SPT-CL J2022-6323	J202209.82-632349.3	305.5409	−63.3970	0.3736 ± 0.0001	rvsao-em	[O II]
SPT-CL J2032-5627	J203214.04-562612.4	308.0585	−56.4368	0.2844 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2043-5035	J204317.52-503531.2	310.8230	−50.5920	0.7225 ± 0.0005	template	Ca II H&K
SPT-CL J2056-5459	J205653.57-545909.1	314.2232	−54.9859	0.7151 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2058-5608	J205822.28-560847.2	314.5928	−56.1465	0.6061 ± 0.0002	rvsao-xc	[O II], Ca II H&K
SPT-CL J2100-4548	J210023.85-454834.6	315.0994	−45.8096	0.7148 ± 0.0002	template	Ca II H&K
SPT-CL J2118-5055	J211853.24-505559.5	319.7218	−50.9332	0.6253 ± 0.0002	template	Ca II H&K
SPT-CL J2124-6124	J212437.81-612427.7	321.1576	−61.4077	0.4375 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J2130-6458	J213056.21-645840.4	322.7342	−64.9779	0.3161 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2135-5726	J213537.41-572630.7	323.9059	−57.4419	0.4305 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2136-6307	J213653.72-630651.5	324.2239	−63.1143	0.9224 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2138-6007	J213800.82-600753.8	324.5034	−60.1316	0.3212 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2145-5644	J214551.96-564453.5	326.4665	−56.7482	0.4813 ± 0.0003	rvsao-xc	Ca II H&K
SPT-CL J2146-4633	J214635.34-463301.7	326.6472	−46.5505	0.9282 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2146-4846	J214605.93-484653.3	326.5247	−48.7815	0.6177 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J2148-6116	J214838.82-611555.9	327.1617	−61.2655	0.5649 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2155-6048	J215555.46-604902.8	328.9811	−60.8175	0.5419 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J2248-4431	J224843.98-443150.8	342.1833	−44.5308	0.3482 ± 0.0001	rvsao-xc	Ca II H&K
SPT-CL J2300-5331	J230039.69-533111.4	345.1654	−53.5198	0.2630 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2325-4111	J232511.70-411213.7	351.2988	−41.2038	0.3624 ± 0.0003	rvsao-xc	Ca II H&K
SPT-CL J2331-5051	J233151.13-505154.1	352.9631	−50.8650	0.5786 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2332-5358	J233227.48-535828.2	353.1145	−53.9745	0.4041 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2337-5942	J233727.52-594204.8	354.3647	−59.7014	0.7788 ± 0.0002	template	Ca II H&K
SPT-CL J2341-5119	J234112.34-511944.9	355.3015	−51.3291	1.0050 ± 0.0005	template	Ca II H&K
SPT-CL J2342-5411	J234245.89-541106.1	355.6912	−54.1850	1.0808 ± 0.0003	template	Ca II H&K
SPT-CL J2344-4243	J234443.90-424312.1	356.1829	−42.7200	0.5981 ± 0.0008	rvsao-em	[O II]
SPT-CL J2355-5056	J235547.48-505540.5	358.9479	−50.9279	0.3184 ± 0.0002	rvsao-xc	Ca II H&K
SPT-CL J2359-5009	J235942.81-501001.7	359.9284	−50.1671	0.7709 ± 0.0003	rvsao-xc	Ca II H&K

is a preferred measurement for those clusters with fewer than 15 member redshifts.

The cluster redshifts and velocity dispersions are calculated using only galaxies identified as members, where membership is established using iterative 3σ clipping on the velocities (see Section 5.1.3). The center at each iteration of 3σ-clipping is the robust biweight average, and σ is calculated from the robust biweight variance, or the gapper estimator in the case where there are fewer than 15 members. We do not make a hard velocity cut; the initial estimate of σ used in the iterative clipping is determined from the galaxies located within 4000 $\mathrm{km\,s^{-1}}$ of the center (in the rest frame), excluding any galaxy in the tails that lies more than 1000 $\mathrm{km\,s^{-1}}$ away from any other galaxy. This initial step is similar in spirit to membership selection algorithms based on gaps in the data, such as the shifting gapper (Fadda et al., 1996), which is used by Sifón et al. (2012). Figure 6.1 shows the velocity histogram for each cluster, as well as an indication of emission-line objects and our determination of member and non-member galaxies.

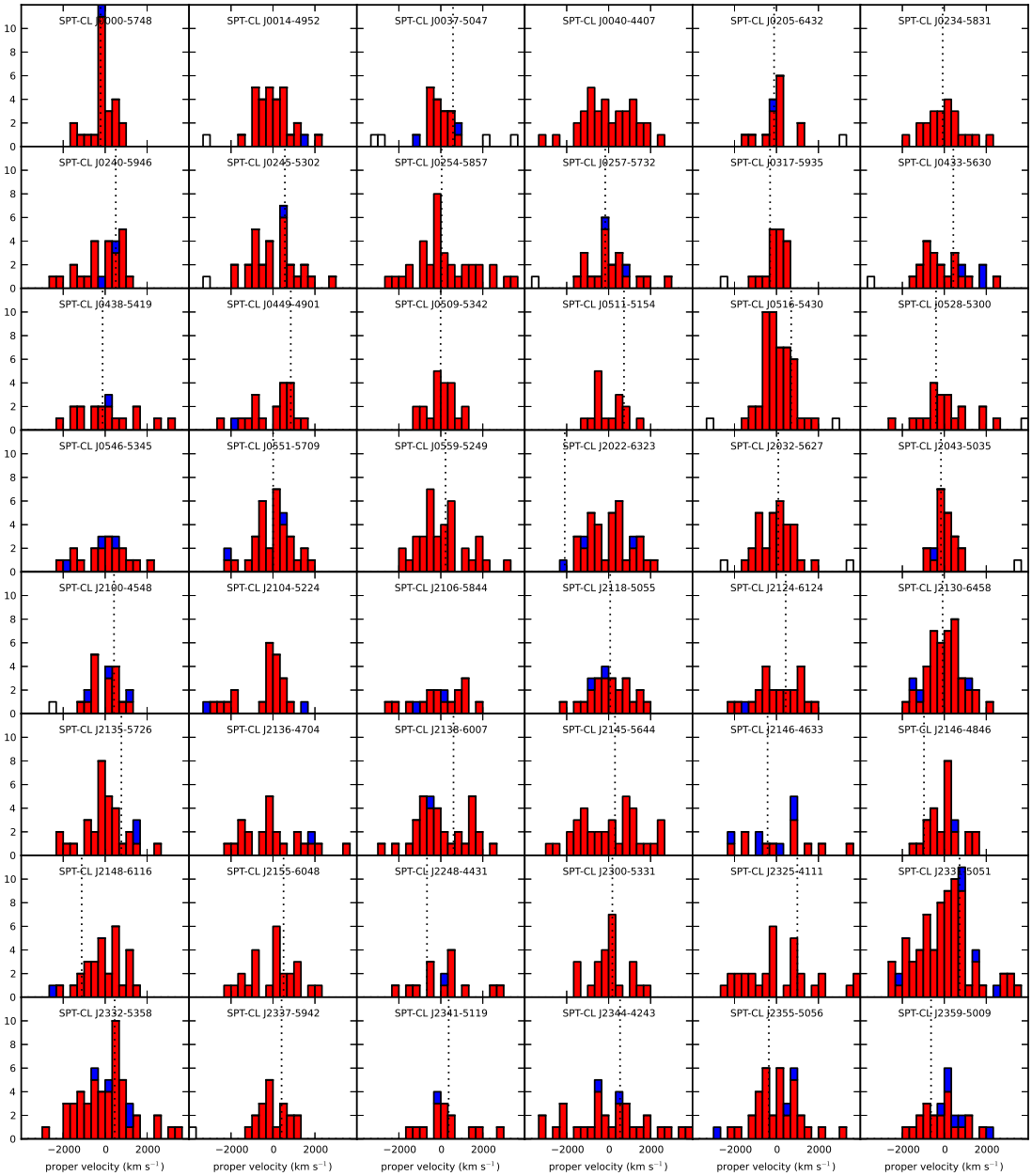

Figure 6.1: Histograms showing the proper velocities of galaxies selected as members (red for passive galaxies, blue for emission-line galaxies), non-members (white) and the central galaxy proper velocity (dotted line, see Section 6.1.3; not measured for six of the clusters, mostly at high redshift).

Some entries in Table 6.2 have a star-shaped flag $\star$ in the SPT ID column, which highlights possibly less reliable dispersion measurements. These are 7 clusters that have fewer than 15 measured member redshifts[1], as well as three others. In one case, SPT-CL J0205-6432, $N_{\rm members} = 15$, the gapper and biweight dispersions differ by more than one sigma[2], and in two cases, SPT-CL J2104-5224 with $N_{\rm members} = 23$, and SPT-CL J2341-5119 with $N_{\rm members} = 15$, the jackknife confidence interval on the biweight variance is very large and can be considered to have failed. We take the statistical failures in these three cases as an indication that the sampling is inadequate.

6.1.3 Central galaxy peculiar velocities

For most of the clusters in this work, that is 49 clusters, we have a spectrum of the central galaxy, which we visually select as a large, bright, typically cD-type galaxy that is close to the SZ center and that appears to be central to the distribution of galaxies; it often, but not always, coincides with the brightest cluster galaxy (BCG; Skibba et al., 2011). Song et al. (2012), whose cluster sample overlaps with the one presented here, select the brightest red galaxy within a projected radius of $R_{200c,\rm SZ}$, which they call the rBCG. Table 6.3 lists coordinates, redshifts, and spatial offsets (from the SZ center) of the objects identified as central galaxies. The cases in which the choice of central galaxy differs from the Song et al. (2012) rBCG are indicated by an asterisk (this difference arises from our criterion of centrality to the spatial distribution of galaxies). In some cases, Song et al. (2012) chose not to designate a

[1]Once again, this is a somewhat arbitrary cutoff. See note at the end of Section 3.1

[2]Since these are not independent measurements but rather two estimates of the same quantity from the same data, we consider a one-sigma discrepancy to be large.

Table 6.2: Cluster redshifts and velocity dispersions
(*This table is continued on next page.*) This table shows the number N ($\equiv N_{\rm members}$) of spectroscopic members as determined by iterative 3σ clipping, the aperture radius a within which they were sampled in units of $R_{200c,\rm SZ}$, the robust biweight average redshift z with the uncertainty in the last two digits in parentheses, the gapper scale $\sigma_{\rm G}$ and the biweight dispersion $\sigma_{\rm BI}$. All confidence intervals are from jackknife resampling. The star flag $\star$ in the SPT ID column indicates less reliable dispersion measurements (see Section 6.1.2) $\sigma_{\rm SZ}$ is the equivalent dispersion from the SPT mass, used in the construction of the stacked cluster. No uncertainty is given as it was not used in our simple analysis.

SPT ID (and flag)	N	a ($R_{200c,\rm SZ}$)	z	$\sigma_{\rm SZ}$ ($\rm km\,s^{-1}$)	$\sigma_{\rm G}$ ($\rm km\,s^{-1}$)	$\sigma_{\rm BI}$ ($\rm km\,s^{-1}$)
SPT-CL J0000-5748	26	1.0	0.7019(05)	937	598 ± 109	563^{+124}_{-161}
SPT-CL J0014-4952	29	1.3	0.7520(08)	1024	812 ± 115	811^{+107}_{-124}
SPT-CL J0037-5047	18	1.6	1.0262(08)	981	550 ± 97	555^{+96}_{-117}
SPT-CL J0040-4407	36	0.4	0.3498(08)	1169	1275 ± 151	1277^{+140}_{-157}
SPT-CL J0118-5156 $\star$	14	0.9	0.7050(27)	874	948 ± 186	986^{+203}_{-257}
SPT-CL J0205-5829	9	1.3	1.3219(07)	1104	-	-
SPT-CL J0205-6432 $\star$	15	1.1	0.7436(04)	871	687 ± 204	340^{+219}_{-340}
SPT-CL J0233-5819 $\star$	10	0.9	0.6635(15)	891	783 ± 183	800^{+165}_{-210}
SPT-CL J0234-5831	22	0.3	0.4149(07)	1079	929 ± 160	926^{+157}_{-190}
SPT-CL J0240-5946	25	0.4	0.4004(10)	950	999 ± 150	1014^{+147}_{-172}
SPT-CL J0245-5302	29	0.4	0.3003(09)	1130	1128 ± 157	1131^{+143}_{-163}
SPT-CL J0254-5857	35	0.4	0.4371(14)	1079	1431 ± 190	1483^{+212}_{-248}
SPT-CL J0257-5732	22	0.6	0.4337(08)	806	1039 ± 189	1024^{+176}_{-213}
SPT-CL J0317-5935	17	0.5	0.4691(04)	839	473 ± 95	473^{+86}_{-105}
SPT-CL J0433-5630	22	0.7	0.6922(13)	824	1084 ± 172	1090^{+158}_{-185}
SPT-CL J0438-5419	18	0.5	0.4223(11)	1196	1428 ± 265	1422^{+237}_{-286}
SPT-CL J0449-4901	20	0.7	0.7898(20)	987	1067 ± 167	1090^{+147}_{-170}
SPT-CL J0509-5342	21	0.8	0.4616(07)	963	670 ± 98	678^{+73}_{-82}
SPT-CL J0511-5154	15	0.9	0.6447(12)	881	778 ± 138	791^{+119}_{-141}
SPT-CL J0516-5430	48	0.4	0.2940(05)	1000	721 ± 76	724^{+73}_{-82}
SPT-CL J0528-5300	20	1.2	0.7693(14)	859	1179 ± 235	1181^{+237}_{-299}
SPT-CL J0533-5005	4	0.4	0.8813(04)	834	-	-
SPT-CL J0534-5937	3	0.4	0.5757(04)	797	-	-
SPT-CL J0546-5345	21	0.8	1.0661(17)	1083	1162 ± 193	1191^{+202}_{-244}
SPT-CL J0551-5709	34	0.6	0.4243(07)	853	962 ± 128	966^{+125}_{-144}
SPT-CL J0559-5249	37	0.8	0.6092(08)	1072	1135 ± 139	1146^{+144}_{-164}

Table 6.2: Cluster redshifts and velocity dispersions (continued)

SPT ID (and flag)	N	a ($R_{200c,\mathrm{SZ}}$)	z	σ_{SZ} ($\mathrm{km\,s^{-1}}$)	σ_{G} ($\mathrm{km\,s^{-1}}$)	σ_{BI} ($\mathrm{km\,s^{-1}}$)
SPT-CL J2022-6323	37	0.4	0.3832(07)	852	1076 ± 103	1080^{+92}_{-100}
SPT-CL J2032-5627	31	0.3	0.2841(05)	900	771 ± 94	777^{+84}_{-94}
SPT-CL J2040-5725	5	0.9	0.9295(32)	894	-	-
SPT-CL J2043-5035	21	1.1	0.7234(06)	975	509 ± 71	524^{+72}_{-83}
SPT-CL J2056-5459 ⋆	12	0.7	0.7185(10)	851	704 ± 206	642^{+233}_{-401}
SPT-CL J2058-5608	9	0.9	0.6065(08)	787	-	-
SPT-CL J2100-4548	19	1.4	0.7124(09)	814	731 ± 102	733^{+89}_{-101}
SPT-CL J2104-5224 ⋆	23	1.5	0.7990(15)	858	1176 ± 211	1153^{+515}_{-1153}
SPT-CL J2106-5844	18	1.0	1.1312(18)	1288	1216 ± 218	1228^{+210}_{-254}
SPT-CL J2118-5055	25	1.2	0.6249(09)	864	981 ± 156	982^{+149}_{-176}
SPT-CL J2124-6124	24	0.6	0.4354(10)	918	1151 ± 149	1153^{+133}_{-151}
SPT-CL J2130-6458	47	0.5	0.3164(05)	887	897 ± 97	903^{+94}_{-105}
SPT-CL J2135-5726	33	0.4	0.4269(08)	980	1020 ± 151	1029^{+163}_{-194}
SPT-CL J2136-4704	24	0.6	0.4247(12)	875	1461 ± 227	1461^{+202}_{-234}
SPT-CL J2136-6307 ⋆	10	0.8	0.9258(23)	889	1244 ± 301	1269^{+273}_{-351}
SPT-CL J2138-6007	34	0.3	0.3185(09)	1017	1269 ± 141	1303^{+132}_{-147}
SPT-CL J2145-5644	37	0.5	0.4798(13)	1029	1634 ± 189	1638^{+170}_{-190}
SPT-CL J2146-4633	17	1.0	0.9309(23)	1061	1558 ± 284	1576^{+250}_{-298}
SPT-CL J2146-4846	26	0.9	0.6230(07)	877	772 ± 111	784^{+104}_{-121}
SPT-CL J2148-6116	30	0.6	0.5707(11)	899	969 ± 139	966^{+130}_{-150}
SPT-CL J2155-6048	25	0.8	0.5393(11)	798	1157 ± 152	1162^{+142}_{-162}
SPT-CL J2248-4431	15	0.2	0.3512(14)	1399	1304 ± 306	1301^{+291}_{-379}
SPT-CL J2300-5331	24	0.3	0.2623(07)	824	887 ± 139	920^{+121}_{-139}
SPT-CL J2301-5546 ⋆	11	0.7	0.7479(22)	856	1242 ± 375	1261^{+414}_{-648}
SPT-CL J2325-4111	33	0.6	0.3579(13)	1051	1926 ± 273	1921^{+255}_{-294}
SPT-CL J2331-5051	78	0.9	0.5748(06)	970	1363 ± 119	1382^{+132}_{-146}
SPT-CL J2332-5358	53	0.6	0.4020(08)	1018	1253 ± 146	1240^{+145}_{-164}
SPT-CL J2337-5942	19	0.9	0.7764(09)	1188	700 ± 99	707^{+95}_{-110}
SPT-CL J2341-5119 ⋆	15	1.1	1.0025(10)	1091	1111 ± 289	959^{+422}_{-959}
SPT-CL J2342-5411 ⋆	11	1.5	1.0746(24)	893	1278 ± 338	1268^{+289}_{-379}
SPT-CL J2344-4243	32	0.7	0.5952(16)	1313	1824 ± 231	1878^{+236}_{-271}
SPT-CL J2347-5158 ⋆	12	0.9	0.8693(07)	786	630 ± 162	635^{+159}_{-215}
SPT-CL J2355-5056	37	0.5	0.3200(07)	858	1124 ± 162	1104^{+149}_{-173}
SPT-CL J2359-5009	26	0.9	0.7747(09)	892	951 ± 135	950^{+123}_{-141}

rBCG due to the presence of bright foreground objects masking part of the cluster, and these are indicated with two asterisks. The results in this section would not change in any significant way if we were to adopt the Song et al. (2012) rBCGs as the central galaxies, which in effect would mean reducing our sample to a subsample.

We calculate the proper velocities of the central galaxies with respect to the cluster average; the uncertainty in that measurement is the quadrature sum of the uncertainty in the cluster velocity (from the cosmological redshift) and in the galaxy velocity. This proper velocity is then compared to the cluster's velocity dispersion $\sigma_{\rm BI}$. Figure 6.2 shows a normalized histogram (in blue diagonal lines) of the $v_p/\sigma_{\rm BI}$ ratio for the central galaxies of the 41 clusters for which the dispersion is calculated from more than 15 members. This distribution is centered around zero (average of 0.04 ± 0.10), as expected.

The observed distribution looks like a continuous peaked distribution, with the exception of perhaps an outlier. The large proper velocities of any outliers could be due to a mis-identification of the central galaxy or poor sampling of the cluster potential by the selected galaxies, or could indicate that the cluster is in a disrupted dynamical state. After inspection of the data in hand, it is not clear which of these explanations account for the offset in the case of our most significant deviation from zero, SPT-CL J2022-6323 (the point seen on the far left of the histogram in Figure 6.2).

In the limiting case where the central galaxy always has a proper velocity of exactly zero kilometers per second, this distribution would still have a nonzero width because of the large proper velocity errors. This zero-velocity distribution is over-plotted as a

dashed line in Figure 6.2 for reference. It was generated by sampling randomly from Gaussians of mean zero and standard deviation given by the uncertainties in $v_p/\sigma_{\rm BI}$. A KS test shows that the distribution of central galaxy velocities is significantly different from this zero-velocity distribution: the KS-test p-value for the null hypothesis of the two distributions being drawn from the same parent distribution is 0.005, whether the possible outlier is included or not. It has been found at low redshift that many BCGs have a proper velocity significantly different from zero; for example, Coziol et al. (2009) found that the BCGs of a sample of 452 Abell clusters have a median peculiar velocity that is 32% of their clusters' radial velocity dispersion. Including a correction for our large uncertainties, we measure this median to be 28^{+5}_{-9}% in our smaller SZ-selected sample, which is consistent with their measurement.

It is expected that the most massive cluster galaxy will gradually come to rest at the center of the potential through dynamical friction, so that this distribution should be narrow. Our observed distribution is narrower than that of all member galaxies (white histogram on Figure 6.2). The KS-test p-value for the two distributions being drawn from the same parent distribution is 0.18 for all central galaxies, and 0.10 when the most significant deviation is excluded; in other words, the distributions are only measured to be different with a low statistical significance by that metric.

One of the motivations for looking at the central-galaxy peculiar velocities and spatial offsets from the SZ center was to use them as an indicator of a disturbed dynamical state. Sifón et al. (2012), with a smaller sample of 16 clusters, use the central galaxy proper velocity and spatial offset as two of three criteria to flag "disturbed" clusters, but do not find that they are different from the clusters flagged as "relaxed"

in the comparison of the dynamical mass and SZ properties. Spatial offsets of SPT cluster rBCGs are discussed in Song et al. (2012), where they are shown to be consistent with the spatial offsets with respect to the X-ray peak in X-ray-selected clusters, when the BCG identification is similar (see, e.g., Lin & Mohr, 2004). Haarsma et al. (2010) find in an X-ray-selected sample that except possibly for a handful of very disturbed systems, correlation between spatial BCG offsets and dynamical state of the gas is difficult to quantify. From a velocity offset perspective, our results suggest that it will not be possible to divide SPT clusters along binary dynamical classes using peculiar velocities, since these are consistent with being normally distributed, per the Anderson-Darling test. Also, the velocity and spatial offsets are not correlated and plotting one against the another does not reveal new outliers or new structure.

Table 6.3: Dynamical properties of central galaxies
(*This table is continued on next page.*) Coordinates, peculiar velocity v_p, ratio v_p/σ of peculiar velocity to velocity dispersion, and spatial offsets of central galaxies (the central galaxy selection is explained in Section 6.1.3). The peculiar velocity is the rest-frame velocity with the rest frame defined by the average redshift of the cluster velocities, and the uncertainty presented here includes the uncertainties on both the galaxy redshift and the cluster redshift. The offsets Δr are measured with respect to the SZ center. * indicates that the choice of central galaxy differs from Song et al. (2012), and ** are the cases where Song et al. (2012) chose not to select a rBCG due to the presence of bright foreground objects masking part of the cluster. The v_p/σ in parentheses are more precisely v_p/σ_G and are not used in Figure 6.2 because the dispersion was measured with too few galaxy redshifts; the other values use the biweight velocity dispersion.

SPT ID	Central gal. R.A. (J2000 deg.)	Central gal. Dec. (J2000 deg.)	v_p (km s^{-1})	v_p/σ	Δr_{ang} (arcsec)	Δr_{phys} (kpc)
SPT-CL J0000-5748	0.2502	−57.8093	-219 ± 115	-0.39 ± 0.34	25 ± 4	178 ± 28
SPT-CL J0037-5047	9.4471	−50.7886	582 ± 127	1.05 ± 0.68	28 ± 2	226 ± 18
SPT-CL J0118-5156	19.6032	−51.9413	-505 ± 501	(-0.53 ± 0.63)	23 ± 3	166 ± 19
SPT-CL J0205-5829	31.4511	−58.4801	-16 ± 115	(-0.03 ± 0.19)	24 ± 1	206 ± 12
SPT-CL J0205-6432*	31.2829	−64.5409	-114 ± 77	(-0.17 ± 0.17)	20 ± 5	147 ± 37
SPT-CL J0233-5819	38.2537	−58.3270	-631 ± 271	(-0.81 ± 0.65)	5 ± 11	33 ± 80
SPT-CL J0234-5831	38.6761	−58.5236	-77 ± 162	-0.08 ± 0.18	9 ± 2	48 ± 11
SPT-CL J0240-5946	40.1600	−59.7635	507 ± 224	0.50 ± 0.36	25 ± 5	135 ± 27
SPT-CL J0245-5302	41.3534	−53.0293	567 ± 201	0.50 ± 0.31	58 ± 3	261 ± 14
SPT-CL J0254-5857*	43.5645	−58.9530	40 ± 301	0.03 ± 0.20	16 ± 6	90 ± 36
SPT-CL J0257-5732	44.3372	−57.5483	-163 ± 170	-0.16 ± 0.19	35 ± 3	200 ± 16
SPT-CL J0317-5935	49.3158	−59.5915	-304 ± 100	-0.64 ± 0.46	23 ± 3	136 ± 15
SPT-CL J0433-5630	68.2541	−56.5194	430 ± 234	0.39 ± 0.31	56 ± 20	403 ± 141
SPT-CL J0438-5419	69.5734	−54.3224	-122 ± 250	-0.09 ± 0.18	17 ± 1	94 ± 8
SPT-CL J0449-4901	72.2668	−49.0275	849 ± 337	0.78 ± 0.52	32 ± 5	243 ± 39
SPT-CL J0509-5342	77.3393	−53.7036	-9 ± 160	-0.01 ± 0.24	8 ± 5	45 ± 27
SPT-CL J0511-5154**	77.9288	−51.9102	732 ± 238	0.93 ± 0.61	28 ± 5	197 ± 37
SPT-CL J0516-5430	79.1557	−54.5004	701 ± 152	0.97 ± 0.49	26 ± 5	116 ± 21
SPT-CL J0528-5300	82.0221	−52.9982	-394 ± 243	-0.33 ± 0.30	12 ± 3	93 ± 21
SPT-CL J0551-5709	87.8983	−57.1412	10 ± 155	0.01 ± 0.16	55 ± 10	310 ± 55
SPT-CL J0559-5249	89.9301	−52.8241	225 ± 168	0.20 ± 0.18	15 ± 6	100 ± 44

Table 6.3: Dynamical properties of central galaxies (continued)

SPT ID	Central gal. R.A. (J2000 deg.)	Central gal. Dec. (J2000 deg.)	v_p ($\mathrm{km\,s^{-1}}$)	v_p/σ	Δr_ang (arcsec)	Δr_phys (kpc)
SPT-CL J2022-6323	305.5411	−63.3971	-2092 ± 173	-1.94 ± 0.83	28 ± 5	149 ± 24
SPT-CL J2032-5627	308.0586	−56.4368	88 ± 147	0.11 ± 0.20	80 ± 13	345 ± 56
SPT-CL J2043-5035**	310.8233	−50.5924	-159 ± 203	-0.30 ± 0.42	12 ± 4	88 ± 28
SPT-CL J2056-5459	314.2231	−54.9859	-586 ± 176	(-0.83 ± 0.69)	14 ± 2	99 ± 18
SPT-CL J2058-5608	314.5927	−56.1464	-71 ± 160	(-0.07 ± 0.17)	8 ± 3	53 ± 20
SPT-CL J2100-4548**	315.0996	−45.8096	422 ± 164	0.58 ± 0.37	21 ± 6	149 ± 45
SPT-CL J2118-5055**	319.7218	−50.9333	73 ± 178	0.07 ± 0.19	17 ± 3	114 ± 18
SPT-CL J2124-6124	321.1576	−61.4076	440 ± 209	0.38 ± 0.26	28 ± 7	157 ± 42
SPT-CL J2130-6458	322.7344	−64.9779	-79 ± 138	-0.09 ± 0.16	11 ± 8	49 ± 37
SPT-CL J2135-5726	323.9059	−57.4418	772 ± 181	0.75 ± 0.47	19 ± 3	108 ± 16
SPT-CL J2136-6307	324.2235	−63.1143	-526 ± 357	(-0.42 ± 0.41)	36 ± 7	287 ± 57
SPT-CL J2138-6007	324.5034	−60.1316	605 ± 228	0.46 ± 0.28	5 ± 4	25 ± 17
SPT-CL J2145-5644*	326.4665	−56.7482	303 ± 299	0.18 ± 0.20	6 ± 2	37 ± 15
SPT-CL J2146-4633**	326.6474	−46.5504	-424 ± 368	-0.27 ± 0.28	1 ± 3	4 ± 25
SPT-CL J2146-4846	326.5247	−48.7815	-969 ± 132	-1.24 ± 0.68	28 ± 15	190 ± 105
SPT-CL J2148-6116	327.1617	−61.2655	-1118 ± 220	-1.16 ± 0.66	58 ± 10	382 ± 68
SPT-CL J2155-6048	328.9812	−60.8175	505 ± 209	0.43 ± 0.29	38 ± 11	243 ± 73
SPT-CL J2248-4431	342.1833	−44.5308	-659 ± 313	-0.51 ± 0.43	15 ± 1	74 ± 5
SPT-CL J2300-5331	345.1655	−53.5200	177 ± 200	0.19 ± 0.24	26 ± 3	105 ± 12
SPT-CL J2325-4111	351.2988	−41.2038	994 ± 313	0.52 ± 0.32	36 ± 7	183 ± 37
SPT-CL J2331-5051	352.9631	−50.8650	727 ± 138	0.53 ± 0.26	11 ± 2	75 ± 12
SPT-CL J2332-5358	353.1146	−53.9745	453 ± 177	0.37 ± 0.23	23 ± 12	123 ± 67
SPT-CL J2337-5942	354.3651	−59.7013	410 ± 165	0.58 ± 0.39	24 ± 1	180 ± 8
SPT-CL J2341-5119	355.3015	−51.3291	375 ± 208	(0.34 ± 0.31)	14 ± 5	116 ± 38
SPT-CL J2342-5411	355.6913	−54.1848	897 ± 358	(0.70 ± 0.58)	14 ± 5	117 ± 40
SPT-CL J2344-4243	356.1829	−42.7200	544 ± 441	0.29 ± 0.28	11 ± 1	73 ± 7
SPT-CL J2355-5056	358.9477	−50.9279	-363 ± 181	-0.33 ± 0.24	36 ± 8	166 ± 36
SPT-CL J2359-5009	359.9286	−50.1672	-632 ± 185	-0.67 ± 0.40	32 ± 12	237 ± 89

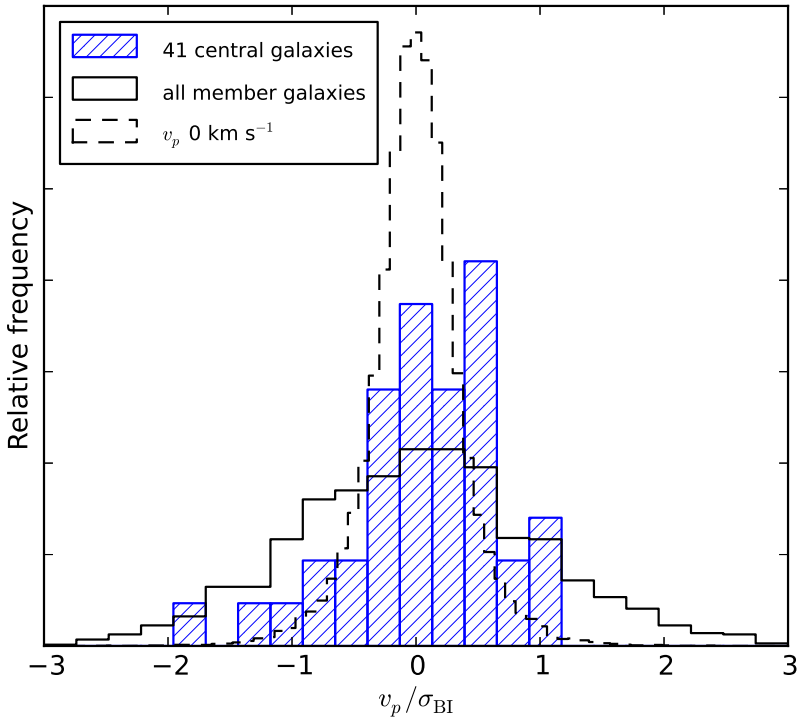

Figure 6.2: Ratio of galaxy proper velocity to cluster velocity dispersion, shown as a normalized histogram both for the central galaxies with well-measured dispersions (blue, hatched) and all member galaxies (white). In the limit where the central galaxy always has a proper velocity of 0 km s^{-1}, the distribution would retain a certain width because of the large errors, and the expected shape of this zero-velocity distribution is overplotted as a dashed line for reference.

6.1.4 Data in the literature: summary and comparison

Table 6.4 contains spectroscopic redshifts and velocity dispersions from the literature for clusters detected by SPT, most notably from Sifón et al. (2012). We independently obtained data for five of these clusters; all of the reported cluster redshifts and dispersions are consistent, and the RMS residuals agree with the size of our uncertainties. This comparison is of particular interest in judging our follow-up

strategy, because the typical number of SPT-reported member galaxies per dispersion is 25 (for $N_{\rm members} \geq 15$), while for the overlapping Sifón et al. (2012) sample it is 55.

We have not had an opportunity to evaluate the extent of overlap between the selected galaxies, which would make the errors correlated, as the data from Sifón et al. (2012) are not available to us at the time of writing. We do know, from that article, that their 48 members for SPT-CL J0546-5345 include our 21 redshifts that were published in Brodwin et al. (2010), and that the redshifts measured in common agree within two sigma.

6.2 Comparison of velocity dispersions with other observables

In this section, we plot the data and make simple fits in an exploratory manner to see how they compare with expectations. Using these fits and comparisons to precisely inform the SZ mass calibration and cosmology will require a different study, where the SZ selection is taken into account and the systematics of our velocity dispersion measurements are fully characterized.

6.2.1 Comparison with SZ-based masses

Figure 6.3 shows the cluster biweight velocity dispersions, from Tables 6.2 and 6.4, plotted against the masses derived from their SPT SZ signal (combined with X-ray observations where applicable; Table 3.1, Section 3.2). The clusters that are included are those with $N_{\rm members} \geq 15$ and $z \geq 0.3$, except for the three flagged for

Table 6.4: Cluster redshifts and velocity dispersions from the literature
Number of member-galaxy redshifts N ($\equiv N_{\rm members}$), cluster redshift and velocity dispersion for clusters found in the literature that are also SPT detections. In the cases where we are presenting our own spectroscopic observations, some of the information from Table 6.2 is repeated in the left half of the present table for reference. References: (1) Sifón et al. (2012); (2) Girardi et al. (1996); (3) Struble & Rood (1999, this paper does not contain confidence intervals); (4) Barrena et al. (2002); (5) Katgert et al. (1998); (6) Buckley-Geer et al. (2011).

SPT / this work				**Literature**				
SPT ID	N	z	$\sigma_{\rm BI}$ ($\rm km\,s^{-1}$)	Lit. ID	N	z	$\sigma_{\rm BI}$ ($\rm km\,s^{-1}$)	Ref.
$z < 0.3$								
SPT-CL J0235-5120	-	-	-	ACT-CL J0235-5121	82	0.2777 ± 0.0001	1063 ± 101	1
SPT-CL J0328-5541	-	-	-	A3126	38	0.0844	1041	3
SPT-CL J0431-6126	-	-	-	A3266	132	0.0594 ± 0.0003	1182^{+100}_{-85}	2
SPT-CL J0658-5556	-	-	-	1E0657-56	71	0.2958 ± 0.0003	1249^{+109}_{-100}	4
SPT-CL J2012-5649	-	-	-	A3667	123	0.0550 ± 0.0004	1208^{+95}_{-84}	2
SPT-CL J2201-5956	-	-	-	A3827	22	0.0983 ± 0.0010	1103^{+252}_{-138}	5
$z \geq 0.3$								
SPT-CL J0102-4915	-	-	-	ACT-CL J0102-4915	89	0.8701 ± 0.0001	1321 ± 106	1
SPT-CL J0232-5257	-	-	-	ACT-CL J0232-5257	64	0.5560 ± 0.0001	884 ± 110	1
SPT-CL J0236-4937	-	-	-	ACT-CL J0237-4939	65	0.3344 ± 0.0001	1280 ± 89	1
SPT-CL J0304-4921	-	-	-	ACT-CL J0304-4921	71	0.3922 ± 0.0001	1109 ± 89	1
SPT-CL J0330-5227	-	-	-	ACT-CL J0330-5227	71	0.4417 ± 0.0001	1238 ± 98	1
SPT-CL J0346-5438	-	-	-	ACT-CL J0346-5438	88	0.5297 ± 0.0001	1075 ± 74	1
SPT-CL J0438-5419	18	0.4223 ± 0.0011	1422^{+237}_{-286}	ACT-CL J0438-5419	65	0.4214 ± 0.0001	1324 ± 105	1
SPT-CL J0509-5342	21	0.4616 ± 0.0007	678^{+73}_{-82}	ACT-CL J0509-5341	76	0.4607 ± 0.0001	846 ± 111	1
SPT-CL J0521-5104	-	-	-	ACT-CL J0521-5104	24	0.6755 ± 0.0003	1150 ± 163	1
SPT-CL J0528-5300	20	0.7693 ± 0.0014	1181^{+237}_{-299}	ACT-CL J0528-5259	55	0.7678 ± 0.0001	928 ± 111	1
SPT-CL J0546-5345	21	1.0661 ± 0.0017	1191^{+202}_{-244}	ACT-CL J0546-5345	48	1.0663 ± 0.0002	1082 ± 187	1
SPT-CL J0559-5249	37	0.6092 ± 0.0008	1146^{+144}_{-164}	ACT-CL J0559-5249	31	0.6091 ± 0.0003	1219 ± 118	1
SPT-CL J0616-5227	-	-	-	ACT-CL J0616-5227	18	0.6838 ± 0.0004	1124 ± 165	1
SPT-CL J2351-5452	-	-	-	SCSOJ235138-545253	30	0.3838 ± 0.0008	855^{+108}_{-96}	6

statistical indications of unreliable dispersions in Table 6.2. The scaling relationship between velocity dispersion and mass from Saro et al. (2012) (see Section 1.3.1) based on N-body simulations is over-plotted as a solid line.

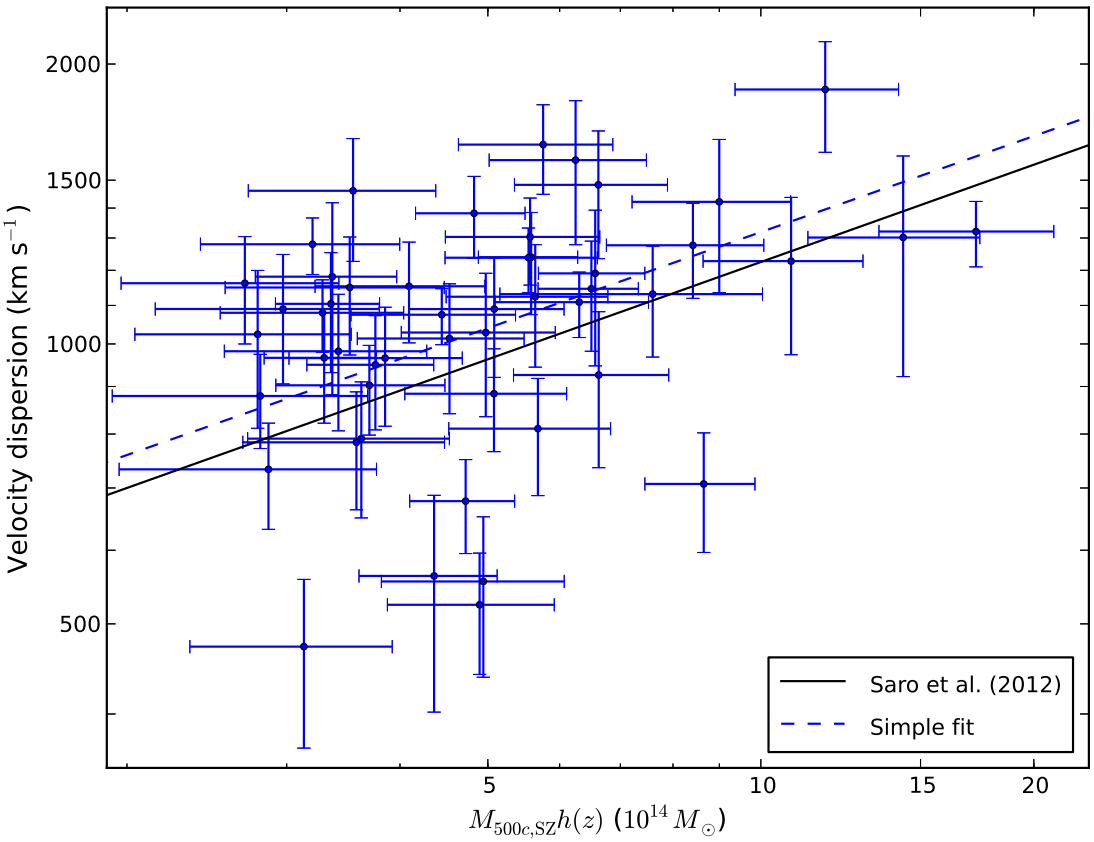

Figure 6.3: Cluster biweight velocity dispersions from Tables 6.2 and 6.4 as a function of SPT SZ-based masses (Table 3.1, Section 3.2) for clusters with $N_{\mathrm{members}} \geq 15$ and $z \geq 0.3$. The figure also shows the scaling relationship predicted from numerical models as a solid line (Saro et al., 2012).

The velocity dispersions appear to have a systematic offset from the model prediction; although we frame the following in terms of the velocity dispersion normalization, we note that any offset would be a relative one, and could be due to a combination of systematics in the measurement of the dispersion and of $M_{500c,\mathrm{SZ}}$. We carry out a simple fit where we fix the slope of the dispersion-$M_{200c,\mathrm{SZ}}$ relationship[3] to that of the Saro et al. (2012) scaling relation, 0.343, and fit for the overall normalization. The resulting velocity dispersion normalization, shown by the dashed blue line, is $(10 \pm 4)\%$ higher than the simulated relation. The size of this normalization offset is consistent with the expected size of systematic biases, discussed in Chapter 7, in particular with the bias expected from the mock observations of simulated clusters from Saro et al. (2012).

The measured scatter in $\ln \sigma$ at fixed mass is $(28 \pm 3)\%$. According to the Anderson-Darling test, the residuals are significantly non-Gaussian at a 97.5% confidence level, but this significance diminishes gradually and ultimately falls below 85% confidence when the calculation is repeated with the one, then two, then three points with the lowest dispersions are excluded (respectively SPT-CL J0317-5935, SPT-CL J2043-5035, and SPT-CL J0037-5047), which suggests normal scatter in $\ln \sigma$ – or log-normal in dispersion –, with a heavy low tail. Sources of scatter relating to the velocity dispersion contributing to the measured scatter are the intrinsic scatter of the dispersion – mass relation (12%), the statistical errors on the dispersion measurements (the mean statistical uncertainty in dispersion is 15%), and extra scatter from systematic effects. Those systematic effects can be related to the membership

[3]Figure 6.3 shows $M_{500c,\mathrm{SZ}}$, but the scaling relation from N-body simulations exists for M_{200c}.

selection, spatial sampling of the cluster, to the sampling of its luminosity function, and to the effect of interlopers; see the discussion in Section 7. Analysis of mock observations from simulated clusters indicate that the combined measurement scatter (intrinsic, statistical, and systematic) is 26% (Saro et al., 2012), in good agreement with the level of the scatter observed here. The resampling exploration of the 3σ clipping membership selection presented in Section 5.3 implies a scatter floor around 22% to which other systematic effects are added.

The errors in $M_{500c,\mathrm{SZ}}$ will also contribute to the observed scatter. The mean uncertainty in $M_{500c,\mathrm{SZ}}$, which includes the effect of intrinsic scatter, is 21%, which translates to 7% in dispersion, a small contribution when added in quadrature with the other sources just discussed.

6.2.2 Comparison with X-ray observations

Turning to X-ray data gives us the opportunity to compare the velocity dispersions and X-ray properties of SPT clusters with existing data on comparable systems, albeit at lower redshift. We use X-ray observations from the SPT *Chandra X-ray Visionary Project* (PI: B. Benson) which is observing the 80 most massive SPT-selected clusters with *Chandra*. This mass-limited sample has been observed and reduced in a uniform fashion, as outlined in Benson et al. (2013; integrated quantities) and McDonald et al. (2013; X-ray profiles). Slight deviations from previously-published values for some clusters (Andersson et al., 2011) in the characteristic ICM temperature, T_X, and Y_X-derived mass, M_{500c,Y_X}, are due to improvements in the X-ray reduction and analysis pipeline (e.g., centroiding, substructure masking, Galactic/extragalactic background

modeling, updated calibration, etc).

From the literature, we take the X-ray temperatures and $M_{500c,\mathrm{Y_X}}$, of the low-z sample of Vikhlinin et al. (2009a), as these data were reduced and analyzed using the same pipeline. The velocity dispersions for many of those galaxy clusters were calculated in a uniform way in Girardi et al. (1996). These velocity dispersion measurements were made with a different galaxy selection and more cluster members, and so will carry different systematics from our own.

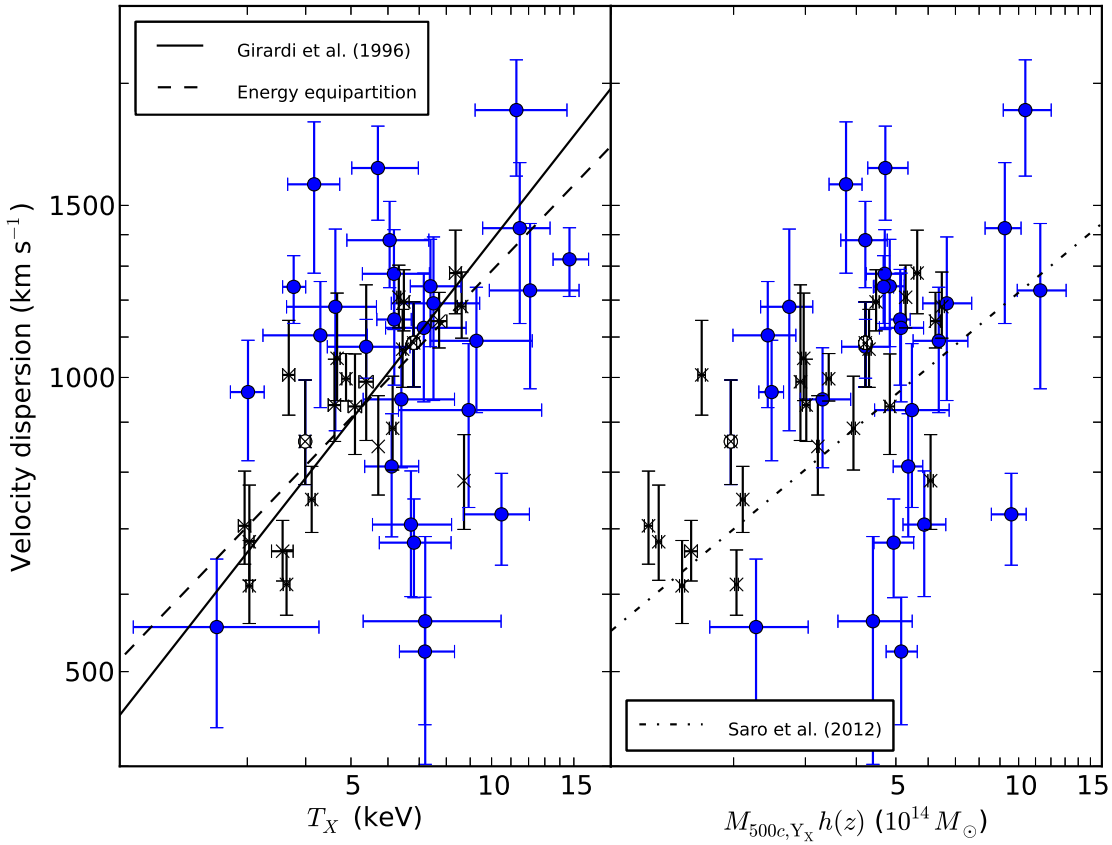

Figure 6.4: Velocity dispersion compared to X-ray properties. The blue points are our sample, and the black crosses are the data from the literature, with X-ray data from Vikhlinin et al. (2009a) and dispersions from Girardi et al. (1996); two of them are also low-redshift SPT detections and are circled. *Left panel:* velocity dispersion vs. X-ray temperature. The solid line is the best-fit scaling relation from Girardi et al. (1996). The dashed line shows the scaling expected if galaxies and gas were both in equilibrium with the gravitational potential. *Right panel:* velocity dispersion vs. M_{500c,Y_X}. The dot-dashed line is the N-body scaling relation from (Saro et al., 2012).

Table 6.5: Literature X-ray and velocity dispersion data
(This table is continued on the next page.) SPT follow-up X-ray data and data from the literature used in Figure 6.4. For the SPT data, the redshift, number of member-galaxy redshifts N ($\equiv N_{\rm members}$) and velocity dispersion from Tables 6.2 and 6.4 are repeated for reference. The literature clusters draw their velocity dispersion from Girardi et al. (1996) and X-ray properties from Vikhlinin et al. (2009a).

Cluster ID	z	N	$\sigma_{\rm BI}$ ($\rm km\,s^{-1}$)	T_X (keV)	M_{500c,Y_X} ($10^{14} M_\odot$)
SPT-CL					
J0000-5748	0.702	26	563^{+124}_{-161}	$7.21^{+3.27}_{-1.90}$	$4.35^{+1.08}_{-0.78}$
J0014-4952	0.752	29	811^{+107}_{-124}	$6.11^{+0.87}_{-0.76}$	$5.16^{+0.44}_{-0.42}$
J0037-5047	1.026	18	555^{+96}_{-117}	$2.58^{+1.69}_{-0.87}$	$1.87^{+0.64}_{-0.43}$
J0040-4407	0.350	36	1277^{+140}_{-157}	$6.18^{+1.18}_{-0.89}$	$5.66^{+0.66}_{-0.56}$
J0102-4915	0.870	89	1321 ± 106	$14.68^{+1.45}_{-1.14}$	$16.41^{+0.99}_{-0.84}$
J0234-5831	0.415	22	926^{+157}_{-190}	$8.93^{+3.87}_{-2.61}$	$6.37^{+1.49}_{-1.21}$
J0330-5227	0.442	71	1238 ± 98	$3.77^{+0.23}_{-0.20}$	$5.35^{+0.31}_{-0.30}$
J0346-5438	0.530	88	1075 ± 74	$5.39^{+1.30}_{-0.94}$	$4.60^{+0.68}_{-0.58}$
J0438-5419	0.422	18	1422^{+237}_{-286}	$11.49^{+1.85}_{-1.92}$	$10.72^{+1.03}_{-1.13}$
J0449-4901	0.790	20	1090^{+147}_{-170}	$9.28^{+2.94}_{-1.89}$	$6.00^{+1.07}_{-0.81}$
J0509-5342	0.462	21	678^{+73}_{-82}	$6.82^{+1.37}_{-1.07}$	$5.60^{+0.67}_{-0.59}$
J0516-5430	0.294	48	724^{+73}_{-82}	$10.50^{+1.55}_{-1.80}$	$11.91^{+1.04}_{-1.27}$
J0528-5300	0.769	20	1181^{+237}_{-299}	$4.63^{+1.04}_{-0.99}$	$2.61^{+0.37}_{-0.39}$
J0546-5345	1.066	21	1191^{+202}_{-244}	$7.51^{+1.92}_{-1.41}$	$5.37^{+0.81}_{-0.68}$
J0551-5709	0.424	34	966^{+125}_{-144}	$3.01^{+0.25}_{-0.25}$	$2.87^{+0.20}_{-0.20}$
J0559-5249	0.609	37	1146^{+144}_{-164}	$6.19^{+0.54}_{-0.51}$	$5.35^{+0.31}_{-0.31}$
J0616-5227	0.684	18	1124 ± 165	$7.16^{+1.66}_{-1.23}$	$5.15^{+0.70}_{-0.59}$
J2043-5035	0.723	21	524^{+72}_{-83}	$7.20^{+1.13}_{-0.85}$	$5.04^{+0.48}_{-0.41}$
J2106-5844	1.131	18	1228^{+210}_{-254}	$12.09^{+3.30}_{-2.20}$	$8.78^{+1.38}_{-1.07}$
J2145-5644	0.480	37	1638^{+170}_{-190}	$5.71^{+1.26}_{-0.69}$	$5.29^{+0.72}_{-0.50}$
J2146-4633	0.931	17	1576^{+250}_{-298}	$4.17^{+0.56}_{-0.51}$	$3.28^{+0.31}_{-0.30}$
J2331-5051	0.575	78	1382^{+132}_{-146}	$6.05^{+1.29}_{-1.15}$	$4.48^{+0.59}_{-0.58}$
J2332-5358	0.402	53	1240^{+145}_{-164}	$7.40^{+1.20}_{-0.70}$	$5.66^{+0.48}_{-0.48}$
J2337-5942	0.776	19	707^{+95}_{-110}	$6.72^{+1.49}_{-1.16}$	$5.57^{+0.72}_{-0.63}$
J2344-4243	0.595	32	1878^{+236}_{-271}	$11.30^{+3.20}_{-2.08}$	$10.93^{+1.72}_{-1.28}$
J2355-5056	0.320	37	1104^{+149}_{-173}	$4.30^{+1.12}_{-1.06}$	$2.97^{+0.51}_{-0.53}$
J2359-5009	0.775	26	950^{+123}_{-141}	$6.41^{+1.92}_{-1.10}$	$3.14^{+0.54}_{-0.37}$

Table 6.5: Literature X-ray and velocity dispersion data (continued)

Cluster ID	z	N	$\sigma_{\rm BI}$ ($\rm km\,s^{-1}$)	T_X (keV)	$M_{500c,\rm Y_X}$ ($10^{14} M_\odot$)
Literature					
A3571	0.039	70	1085^{+110}_{-107}	6.81 ± 0.10	5.90 ± 0.06
A2199	0.030	51	860^{+134}_{-83}	3.99 ± 0.10	2.77 ± 0.05
A496	0.033	151	750^{+61}_{-56}	4.12 ± 0.07	2.96 ± 0.04
A3667	0.056	123	1208^{+95}_{-84}	6.33 ± 0.06	7.35 ± 0.07
A754	0.054	83	784^{+90}_{-85}	8.73 ± 0.00	8.47 ± 0.13
A85	0.056	131	1069^{+105}_{-92}	6.45 ± 0.10	5.98 ± 0.07
A1795	0.062	87	887^{+116}_{-83}	6.14 ± 0.10	5.46 ± 0.06
A3558	0.047	206	997^{+61}_{-51}	4.88 ± 0.10	4.78 ± 0.07
A2256	0.058	47	1279^{+136}_{-117}	8.37 ± 0.24	7.84 ± 0.15
A3266	0.060	132	1182^{+100}_{-85}	8.63 ± 0.18	9.00 ± 0.13
A401	0.074	123	1142^{+80}_{-70}	7.72 ± 0.30	8.63 ± 0.24
A2052	0.035	62	679^{+97}_{-59}	3.03 ± 0.07	1.84 ± 0.03
Hydra-A	0.055	82	614^{+52}_{-43}	3.64 ± 0.06	2.83 ± 0.03
A119	0.044	80	850^{+108}_{-92}	5.72 ± 0.00	4.50 ± 0.03
A2063	0.034	91	664^{+50}_{-45}	3.57 ± 0.19	2.21 ± 0.08
A1644	0.048	92	937^{+107}_{-77}	4.61 ± 0.14	4.21 ± 0.09
A3158	0.058	35	1046^{+174}_{-99}	4.67 ± 0.07	4.13 ± 0.05
MKW3s	0.045	30	612^{+69}_{-52}	3.03 ± 0.05	2.09 ± 0.03
A3395	0.051	107	934^{+123}_{-100}	5.10 ± 0.17	6.74 ± 0.18
A399	0.071	92	1195^{+94}_{-79}	6.49 ± 0.17	6.18 ± 0.11
A576	0.040	48	1006^{+138}_{-91}	3.68 ± 0.11	2.34 ± 0.05
A2634	0.030	69	705^{+97}_{-61}	2.96 ± 0.09	1.74 ± 0.04
A3391	0.055	55	990^{+254}_{-128}	5.39 ± 0.19	4.06 ± 0.10

Figure 6.4 shows the velocity dispersion versus X-ray temperature and versus $M_{500c,\rm Y_X}$. The blue points are our data, and the black crosses are the data from the literature; these literature data are listed for reference in Table 6.5.

The left panel of Figure 6.4 shows dispersion versus T_X. The empirical best-fit scaling relation from Girardi et al. (1996), where $\sigma \propto T_X^{0.61}$, is plotted as a solid line; this scaling relation is consistent with the Vikhlinin et al. (2009a) temperatures used here, although it was fit using X-ray temperatures from a different source, David et al. (1993). The comparison to the temperature is especially interesting in that there is, to first order, a simple correspondence between temperature and velocity dispersion. Assuming that the galaxies and gas are both in equilibrium with the

potential (see, e.g., Voit, 2005), then $\sigma^2 = k_{\rm B}T_X/(\mu m_p)$, where m_p is the proton mass, and μ the mean molecular weight (we take $\mu = 0.58$; see Girardi et al., 1996). This energy equipartition line is plotted as a dashed line in the left panel of Figure 6.4. Real clusters show a deviation from this simple model, but it offers an interesting theoretical baseline, one independent of data or simulations. This relation implies that the temperature and velocity dispersion have a similar redshift evolution, which is why the quantities in this plot are uncorrected for redshift.

The X-ray Y_X observable, while not independent from T_X, is expected to be significantly less sensitive to cluster mergers than T_X, with simulations predicting Y_X to have both a lower scatter and to be a less biased mass indicator (see, e.g., Kravtsov et al., 2006; Fabjan et al., 2011). For this reason, we also plot the velocity dispersion against $M_{500c,\rm Y_X}$ (times a redshift-evolution factor), in the right panel of Figure 6.4. The dot-dashed line is the scaling relation predicted from the simulation analysis of Saro et al. (2012).

The residuals of the dispersion-$M_{500c,\rm Y_X}$ relation have a measured vertical scatter of $(34 \pm 5)\%$, which is larger than but consistent with the full-sample $(28 \pm 3)\%$. The extra scatter is not expected to come from the $M_{500c,\rm Y_X}$ measurements, which have a mean statistical uncertainty of 14% and intrinsic scatter in mass at fixed Y_X of 7% (Kravtsov et al., 2006), for a total of 16% when added in quadrature, smaller than the 21% $M_{500c,\rm SZ}$ mean uncertainties (which include the intrinsic scatter).

Chapter 7

Conclusion and next steps

We have presented a program for the optical spectroscopic follow-up of SPT clusters, and the first results from the measurement of the clusters' cosmological redshifts and velocity dispersions. Our resampling analysis has shown that the strategy of obtaining the velocities of few members per galaxy cluster works for obtaining unbiased velocity dispersions as long as a proper statistical methodology is used. Our resampling analysis has also provided a basic understanding of the uncertainties associated with our measurements.

Systematics remain that are not well understood yet, and will need to be for these data to be used to their fullest extent, that is for informing the mass calibration of SPT-SZ clusters, and therefore the derived cosmology. In the following, we reflect on the way to approach mass calibration using the velocity dispersions presented above.

Velocity dispersion systematics and mass scaling relation fit

Achieving a mass calibration of clusters using velocity dispersions relies on the halo dispersion vs. mass scaling relation from N-body simulations. Yet, there is a significant difference between the dark-matter halo dispersions in simulations, and the cluster velocity dispersions that are measured from a color- and magnitude-selected sample of galaxies, with geometric constraints as we have seen in Section 3.3.

The systematic effects that could lead to bias in the dispersions relative to the dark matter simulations (or increase the measured scatter) include:

- Color selection and membership selection. This was discussed in some detail in Section 5.1.3. Saro et al. (2012) suggests that the measured velocity dispersions can be biased by a few to 10% by the color selection and membership selection.

- The sampling of the luminosity function and location of the members in color space. The magnitudes are important as a proxy of the galaxy masses, as more massive galaxies are expected to be affected by dynamical friction; the distribution of the central galaxy velocities in Section 6.1.3 is an illustration of this. The spatial and luminosity-function sampling, as can be characterized by the aperture of the observations and the magnitudes of the targeted galaxies, can change the measured dispersion by order 5% to 10% (Zhang et al., 2011). We note that the geometric constraints of slit placement in multislit spectroscopic observations have the effect of selecting targets of varying magnitudes even if the brightest galaxies are prioritized, so that in the few-mask regime where we are operating, there will not be a large difference in magnitude distribution between different masks.

- The spatial sampling of the cluster. The velocity dispersion profile can vary significantly from one cluster to another and the velocity dispersion is often higher but can be lower in the core of the cluster by a significant amount (Mohr et al., 1996; Girardi et al., 1996). Saro et al. (2012) find the velocity dispersion at small aperture to be biased high on average by a few percent to 10%. The bias can grow to be quite large at apertures larger than $1.5R_{200c}$, because of the large interloper fraction. It should only be a few percent in the aperture range of most of our observations.

Those systematics will need to be accounted for in fitting the scaling relations. We outline two different ways in which this could be done.

Simulations

The first approach is to produce simulations that not only encapsulate the correct dynamics of halos, but also reproduce the characteristics of observations, and fit not a dark matter halo dispersion vs. mass scaling relation, but an "observed" dispersion vs. mass scaling relation. The most suitable simulations so far in the literature can be found in Saro et al. (2012), and it is not a coincidence that some of the authors are SPT collaborators.

Saro et al. (2012) identifies subhalos in the simulation as galaxies, and assigns them colors and magnitudes following the "semi-analytic model" (SAM; De Lucia & Blaizot, 2007). Each cluster is then pseudo-observed along many different lines-of-sight, where the observed galaxies in each observation are selected to lie close to the red sequence, and chosen within a given aperture radius, which corresponds to the

field-of-view of the spectrograph. This approach relies on the color painting of both the cluster and field galaxies to be accurate, which is a possible source of systematic effects.

Low-redshift / archival study

The second approach is to quantify systematics using observations of real clusters of galaxies that have good photometry and extensive spectroscopy, such as hundreds of member galaxy redshifts, and large numbers of field galaxies as well. Such datasets exist at lower redshifts than those of the SPT sample.

The galaxy redshifts can be used first, without any kind of color selection, to derive a mass solely from the phase-space location of galaxies, a procedure that can be replicated in dark-matter only simulations. For instance, the caustic technique (see, e.g. Serra et al., 2011) can yield such a M_{200c}. Once this reference mass is computed, color-selected pseudo-observations resembling our observations of SPT clusters can be sampled from each cluster. Comparison to the reference mass can quantify the systematic effects of distance from the cluster center, as well as galaxy color and magnitude selection, and produce an observational prescription for accounting for them.

Ideally, both the path of simulations and of resampling would be pursued in parallel, and the resampling from existing observations would provide feedback and validation for the simulation work.

Additional studies of simulations and observational validation of those results are necessary future steps for the mass calibration. The present sample is too small

to carry out the next step in this validation cycle, which will require the study of clusters that are densely sampled spatially and in their luminosity function, possibly with hundreds of measured member redshifts, specifically in relation to simulations and the systematics listed above.

Bibliography

Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409

Allington-Smith, J., et al. 1994, PASP, 106, 983

Andersson, K., et al. 2011, ApJ, 738, 48

Appenzeller, I., et al. 1998, The Messenger, 94, 1

Barrena, R., Biviano, A., Ramella, M., Falco, E. E., & Seitz, S. 2002, A&A, 386, 816

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32

Benson, B. A., et al. 2013, ApJ, 763, 147

Biviano, A. 2000, in Constructing the Universe with Clusters of Galaxies, arXiv:astro-ph/0010409

Biviano, A., Murante, G., Borgani, S., Diaferio, A., Dolag, K., & Girardi, M. 2006, A&A, 456, 23

Brodwin, M., et al. 2010, ApJ, 721, 90

——. 2011, ApJ, 732, 33

Buckley-Geer, E. J., et al. 2011, ApJ, 742, 48

Carlstrom, J. E., et al. 2011, PASP, 123, 568

Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, ARA&A, 40, 643

Coziol, R., Andernach, H., Caretta, C. A., Alamo-Martínez, K. A., & Tago, E. 2009, AJ, 137, 4795

Danese, L., de Zotti, G., & di Tullio, G. 1980, A&A, 82, 322

David, L. P., Slyz, A., Jones, C., Forman, W., Vrtilek, S. D., & Arnaud, K. A. 1993, ApJ, 412, 479

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

den Hartog, R., & Katgert, P. 1996, MNRAS, 279, 349

Dressler, A., Hare, T., Bigelow, B. C., & Osip, D. J. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

Eddington, A. S. 1913, MNRAS, 73, 359

Efron, B. 1987, Journal of the American Statistical Association, 82, 171

Evrard, A. E., et al. 2008, ApJ, 672, 122

Fabjan, D., Borgani, S., Rasia, E., Bonafede, A., Dolag, K., Murante, G., & Tornatore, L. 2011, MNRAS, 416, 801

Fadda, D., Girardi, M., Giuricin, G., Mardirossian, F., & Mezzetti, M. 1996, ApJ, 473, 670

Foley, R. J., et al. 2011, ApJ, 731, 86

——. 2003, PASP, 115, 1220

Gilbank, D. G., Yee, H. K. C., Ellingson, E., Gladders, M. D., Loh, Y., Barrientos, L. F., & Barkhouse, W. A. 2008, ApJ, 673, 742

Girardi, M., Biviano, A., Giuricin, G., Mardirossian, F., & Mezzetti, M. 1993, ApJ, 404, 38

Girardi, M., Fadda, D., Giuricin, G., Mardirossian, F., Mezzetti, M., & Biviano, A. 1996, ApJ, 457, 61

Haarsma, D. B., et al. 2010, ApJ, 713, 1037

Hand, N., et al. 2012, Physical Review Letters, 109, 041101

High, F. W., et al. 2012, ApJ, 758, 68

——. 2010, ApJ, 723, 1736

Hook, I. M., Jørgensen, I., Allington-Smith, J. R., Davies, R. L., Metcalfe, N., Murowinski, R. G., & Crampton, D. 2004, PASP, 116, 425

Howell, S. B. 2006, Handbook of CCD Astronomy (Cambridge University Press)

Hwang, H. S., & Lee, M. G. 2008, ApJ, 676, 218

Kasun, S. F., & Evrard, A. E. 2005, ApJ, 629, 781

Katgert, P., Mazure, A., den Hartog, R., Adami, C., Biviano, A., & Perea, J. 1998, A&AS, 129, 399

Kelson, D. D. 2003, PASP, 115, 688

Komatsu, E., et al. 2011, ApJS, 192, 18

Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, ApJ, 650, 128

Kurtz, M. J., & Mink, D. J. 1998, PASP, 110, 934

Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Lin, Y.-T., & Mohr, J. J. 2004, ApJ, 617, 879

Mamon, G. A., Biviano, A., & Murante, G. 2010, A&A, 520, A30

Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, MNRAS, 406, 1759

Marriage, T. A., et al. 2011, ApJ, 737, 61

McDonald, M., et al. 2012, Nature, 488, 349

Mohr, J. J., Geller, M. J., Fabricant, D. G., Wegner, G., Thorstensen, J., & Richstone, D. O. 1996, ApJ, 470, 724

Mosteller, F., & Tukey, J. W. 1977, Data analysis and regression. A second course in statistics (Addison-Wesley)

Planck Collaboration, et al. 2011, A&A, 536, A8

Press, W., & Schechter, P. 1974, ApJ, 187, 425

Quintana, H., Carrasco, E. R., & Reisenegger, A. 2000, AJ, 120, 511

Reichardt, C. L., et al. 2013, ApJ, 763, 127

Saro, A., Bazin, G., Mohr, J., & Dolag, K. 2012, ArXiv e-prints, 1203.5708

Serra, A. L., Diaferio, A., Murante, G., & Borgani, S. 2011, MNRAS, 412, 800

Sifón, C., et al. 2012, ArXiv e-prints, 1201.0991

Skibba, R. A., van den Bosch, F. C., Yang, X., More, S., Mo, H., & Fontanot, F. 2011, MNRAS, 410, 417

Song, J., et al. 2012, ApJ, 761, 22

Stalder, B., et al. 2013, ApJ, 763, 93

Staniszewski, Z., et al. 2009, ApJ, 701, 32

Struble, M. F., & Rood, H. J. 1999, ApJS, 125, 35

Sunyaev, R. A., & Zel'dovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, ApJ, 688, 709

Vanderlinde, K., et al. 2010, ApJ, 722, 1180

Vikhlinin, A., et al. 2009a, ApJ, 692, 1033

——. 2009b, ApJ, 692, 1060

Voit, G. M. 2005, Reviews of Modern Physics, 77, 207

Wade, R. A., & Horne, K. 1988, ApJ, 324, 411

White, M., Cohn, J. D., & Smit, R. 2010, MNRAS, 408, 1818

Williamson, R., et al. 2011, ApJ, 738, 139

Yahil, A., & Vidal, N. V. 1977, ApJ, 214, 347

Zhang, Y.-Y., Andernach, H., Caretta, C. A., Reiprich, T. H., Böhringer, H., Puchwein, E., Sijacki, D., & Girardi, M. 2011, A&A, 526, A105